AD_____

AWARD NUMBER: DAMD17-02-1-0691

TITLE: Center for the Evaluation of Biomarkers for the Early Detection of Breast Cancer

PRINCIPAL INVESTIGATOR: Nicole Urban

CONTRACTING ORGANIZATION: Fred Hutchinson Cancer Research Center
Seattle, WA 98109

REPORT DATE: October 2009

TYPE OF REPORT: Final Addendum

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
1 October 2009	Final Addendum	23 Sep 2008 – 22Sep 2009

4. TITLE AND SUBTITLE

Center for the Evaluation of Biomarkers for the Early Detection of Breast Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
DAMD17-02-1-0691

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Nicole Urban

E-Mail: nurban@fhcrc.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Fred Hutchinson Cancer Research Center
Seattle, WA 98109

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Breast Cancer remains a leading cause of death for women in the US despite the popularity of mammography as a preventive tool. At diagnosis, many breast cancers are at an advanced stage of disease, even for women undergoing annual screening, resulting in costly and painful follow-up procedures. It has been shown by current clinical use of marker such as CA-125 and PSA that molecular markers can increase our ability to diagnose early stage tumors. By combining molecular diagnosis with current imaging analysis of breast tissue, we may further reduce the number of deaths as well as the number of women undergoing surgery unnecessarily. The translational goal of this Center of Excellence (CoE) is the development of a panel of serum markers with decision rules for its use to improve the performance of mammography and other breast cancer screening tools. This year we measured our most promising markers in serum and plasma specimens collected from CoE participants. The specimens are divided in two groups: a Panel Development Set (PDS) and a Panel Validation Set (PVS). Markers were measured in both groups but only results from the PDS have been analyzed. The results of this work are presented in the attached report and will be submitted next year in a manuscript for publication. Due to a limited number of candidate serum markers we have continued our efforts to identify new markers in tissue and to find or develop assays for marker evaluation. A manuscript reporting the results of these additional efforts was submitted for publication in September and another paper describing assay development techniques is in draft.

15. SUBJECT TERMS
Early detection, serum markers, screening, mammography, antibody tests, molecular profiling, marker panel, mammography, biomarker evaluation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	61	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

Introduction

Although mammography significantly reduces its toll, breast cancer remains a leading cause of cancer mortality in the U.S. Many breast cancers are advanced at the time of diagnosis, even among women participating in screening. The discovery of molecular markers associated with breast cancer potentially increases our ability to diagnose early stage tumors. The translational goal of this Center of Excellence (CoE) is a panel of serum markers with decision rules for its use to improve the performance of breast cancer screening that includes mammography. The primary aims of this study are: 1) to validate and refine the ability of candidate biomarkers measurable in blood products to predict disease status; 2) to evaluate panels of serum markers for use as an adjunct to mammography, to detect all breast cancer at a highly curable stage; and 3) to identify the molecular signatures of benign, pre-invasive and invasive breast tissue and explore their associations with serum markers in the panel. Several years ago it became apparent that there were not enough candidate markers ready for validation. Since that time we have made an effort to find and prepare new markers for evaluation in the CoE. This report details research accomplishments during the active project period from February, 2004-September, 2009. Although CoE funding began on September 23 2002, the DOD did not grant us human subjects approval until February (Mammography Tumor Registry) and May (Breast Cancer Early Discovery Study) of 2004. In light of this delay we were granted a 2 year extension to complete our work and the project officially ended on September 22, 2009.

Body

During the course of this study, CoE researchers have focused their efforts and made progress in three areas:

- Biomarker Discovery: identification of promising biomarkers, assay development.
- Biomarker Evaluation: evaluation of individual and assay panels for their ability to detect breast cancer.
- Resource Development: tissue and blood repositories for future research involving biomarker discovery and validation.

During the last half of this project investigators have created and utilized a Breast Discovery Set (BDS) used by CoE investigators and collaborators to evaluate new markers and conduct the molecular profiling work described in project Aim 3, and the Panel Development and Validation Sets (PDS, PVS), used to evaluate candidates in a biomarker panel with the aim of improving the sensitivity of breast cancer screening including mammography (see Tasks 10-13 for more detail). Both the BDS and the PVS may be made available to outside investigators who have promising biomarkers or discovery platforms through an RFA mechanism.

Subject recruitment and specimen collection began in Seattle at Swedish Medical Center (SMC) in May, 2004 and continued through September 18, 2009. Mammography data (assessment codes, follow up recommendations, and breast density) in coordination with family history collected on our baseline questionnaire and the GAIL model[1] are used to determine risk of breast cancer (high, elevated or average risk) for women in the mammography cohort[1, 2]. The surgical cohort consists of women scheduled to undergo breast surgery for malignant or nonmalignant conditions. Surgical recruitment at Cedars-Sinai Medical Center in Los Angeles was led by Drs. Beth and Scott Karlan and took place from July, 2005 through April, 2008. The clinical protocols for each site have been standardized as much as possible with shared data collection instruments and a web-enabled data entry system (Seattle Informatics Management System or SIM). A summary of recruitment to date is provided below in **Table 1**.

Table 1. Summary of cumulative study accrual: May 2004-Sep 2009

Study Population	Cumulative Study Recruitment Goals	TOTAL Actual Accrual*
Mammography Cohort -- High/Elevated Risk	600	570
Mammography Cohort – Average Risk	200	53
Mammography Cohort – Biopsy	400	13
Subtotal: Mammography Cohort	***1200***	***636***
SMC Surgical Cohort: Blood and Tissue	175	90
SMC Surgical Cohort: Blood Only	300	296
Subtotal: Surgical Cohort	***475***	***337***
Cedars Surgical Cohort: Blood and Tissue	200	118
Total	**1475**	**1091**
**includes only women who have donated one or more blood specimens and who have completed a baseline questionnaire.*		

All CoE participants are asked to donate annual serum samples. Fresh frozen tissue is collected from surgical participants at the time of their procedure if the pathologist deems it clinically appropriate. Specimen data are linked to extensive epidemiological and clinical data including demographic information, information from GAIL model variables,

Table 2. CoE Specimen Summary

		Number of Collections* (Seattle Only)						
		Blood Products				**Tissue Products**		
Participant Pathology (Most Severe Dx for that patient)	**Participants with available specimens**	**Serum** (~thirteen 1 ml specimens/collection)	**EDTA Plasma** (~four 1 ml specimens/collection)	**ACD Plasma** (~one 4.5 ml specimen/collection)	**ACD Buffy Coat Cells** (~2 specimens/collection)	**Snap Frozen Tissue** (1-5 specimens/tissue site)	**OCT-Embedded Frozen Tissue** (3 tissue blocks/tissue site)	**Formalin-Fixed, Paraffin-Embedded Tissue (FFPE**, MRI Protocol only)
Atypia	5	11	11	7	7	0	0	0
Benign	18	23	23	17	18	0	0	0
Carcinoma Unknown	1	2	2	2	2	0	0	0
In Situ	60	165	160	91	91	1	1	0
Invasive/Infiltrating	290	732	717	398	400	100	97	8
Normal	585	1550	1535	676	677	3	0	0
Total	**959**	**2483**	**2448**	**1191**	**1195**	**104**	**98**	**8**

"Collection" refers to all specimens collected from a single participant on a given day. Some participants have donated specimens 2 or 3 separate times.

pedigree, breast density, BIRADS assessment code and follow-up information from screening mammograms, ER/PR and Her2 status, staging and grade, and clinical follow-up for treatment and recurrence. Table 2 on page 5 summarizes the CoE specimens stored in our repository. We have also collected a number of matched pre- and post-diagnosis serum specimens from healthy controls that went on to develop breast cancer during the course of the study (Tables 3a-b.)

Table 3a. Preclinical blood samples available: Summary of breast cancer cases diagnosed after enrollment in women who have donated specimens.

	Stage								
Study	**0**	**I**	**IIA**	**IIB**	**III**	**IIIA**	**IV**	**Unavail.**	**Total**
MAMMOGRAPHY COHORT- Average, Elevated and High Risk	6	3	1	1					**11**
MAMMOGRAPHY COHORT- biopsy	2	1			1				**4**
COE Subtotal	***8***	***4***	***1***	***1***	***1***	***0***	***0***	***0***	***15***
OTHER TOR STUDIES	27	24	5	4	0	2	1	3	
Total	**35**	**28**	**6**	**5**	**2**	**2**	**1**	**3**	**19**

Table 3b. Time period in which preclinical (pre-diagnosis) blood samples were collected.

Months from blood draw to breast cancer diagnosis	**Number of cases per time period:**	
	COE Mammography Cohort	**Participants in other TOR studies[1]**
0 – 6	18	18
7 – 12	5	12
13 – 18	7	13
19 – 24	5	10
25 – 30	3	5
31 – 36	4	4
37 – 42	3	4
43 – 48	1	3

The TOR Laboratory has purchased or developed assays for 65 promising candidate markers for evaluation in the BDS and an additional 14 potential biomarkers have been identified and tested in the BDS by our collaborators. The BDS is provided to investigators in a blinded fashion; once data are sent back to the TOR laboratory, investigators may be unblinded to case/control status, histologic subtype, stage, and potentially other variables on request. The BDS includes 66 specimens from 66 patients. Characteristics of the BDS are described below in Table 4. Of the 79 biomarkers analyzed in the BDS, 9 achieved greater than 30% sensitivity to breast cancer cases at the 90% specificity level: SPARC[3,4], GDF15 (MIC1, X065), FN1, SPP1 (Osteopontin), IL17, COL1A1, IGFBP2, CTGF, and MMP7 (Figure 1). Appendix A describes these assays in greater detail. COL1A1, IL17 and SPARC appeared lower among cases vs. controls, while the remaining 6 markers were elevated in the cases. To evaluate complementarily of these the most sensitive markers, a logistic regression model was fit using the 6 most sensitive markers. Coefficients from the fitted model were used to form a panel using all 6 markers. The resulting marker panel achieved 61% sensitivity at 90% specificity (Figure 2).

Unpublished Data

Table 4. BDS: Summary of participant characteristics

	Healthy Controls	In-Situ Cases	Invasive Cases
Number of Collections	33	11	22
Number of Patients	33	11	22
Age at Collection (mean,sd)	53.8 (7.9)	55.6 (4.4)	53.2 (10.8)
Collected at Surgery	0 (0%)	0 (0%)	0 (0%)
Current OC Use	2 (6.1%)	0 (0%)	0 (0%)
Current HRT Use	2 (6.1%)	0 (0%)	2 (9.1%)
Stage 1	N/A	1 (9.1%)	5 (22.7%)
Stage 2	N/A	0 (0%)	12 (54.5%)
Stage 3	N/A	0 (0%)	4 (18.2%)
Tested for ER	N/A	10 (90.9%)	22 (100%)
Tested for PR	N/A	10 (90.9%)	22 (100%)
Tested for HER2	N/A	2 (18.2%)	19 (86.4%)
Triple Neg or HER2 Pos	N/A	5 (45.5%)	9 (40.9%)

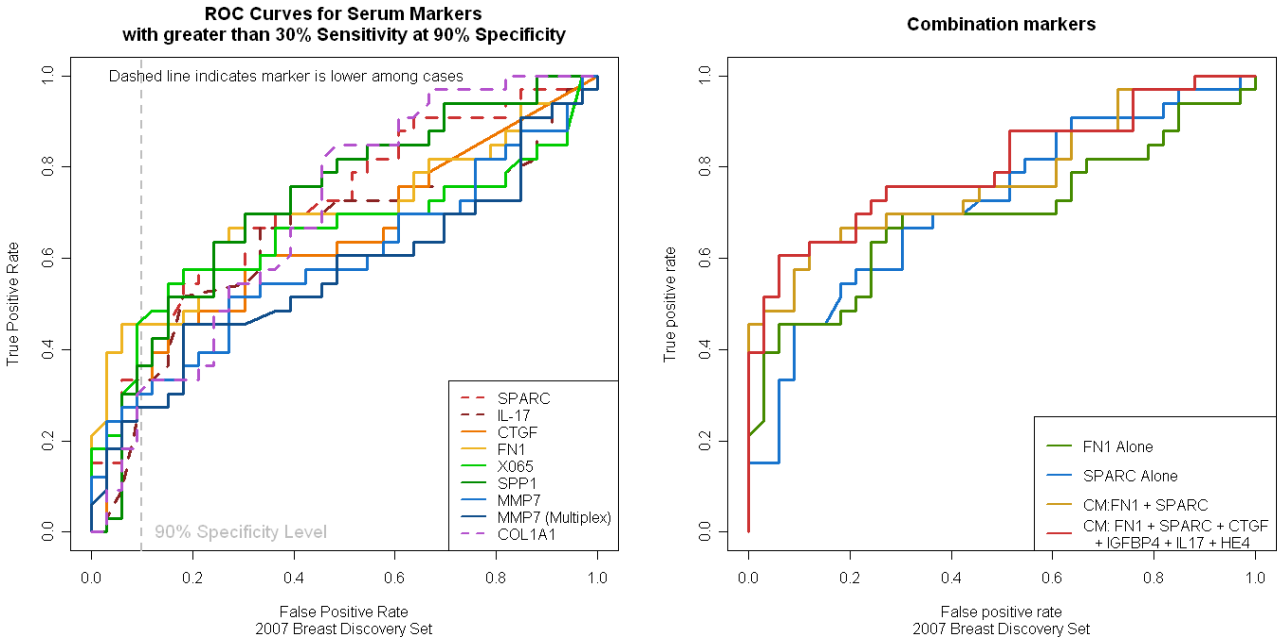

Figure 1. ROC curves for serum markers tested in the BDS with greater than 30% Sensitivity at 90% Specificity. Figure 2. Combined results for 6 best performing markers in the BDS.

This year investigators put together a Breast Panel Validation Set of 1067 specimens collected in Seattle from CoE participants. This large set was divided into two smaller sets: the Panel Development Set (PDS) consisting of 527 specimens from 317 participants, and the Panel Validation Set (PVS) containing 537 specimens from 323 participants. Tables 5a-b below summarize patient characteristics across both sets.

Unpublished Data

Table 5a. Breast Panel Validation Set Patient Characteristics part 1	Healthy Controls		In-Situ Cases		Invasive Cases	
Development (D) or Validation (V) Set	D	V	D	V	D	V
Number of Collections	364	378	24	20	139	139
Number of Patients	204	205	20	17	93	101
Age at Collection	55.7 (8.9)	55.5 (9.6)	57 (9.5)	56.3 (13)	57.9 (12.3)	54.3 (10.6)
Collected at Surgery	0 (0%)	0 (0%)	12 (50%)	11 (55%)	71 (51.1%)	75 (54%)
Current OC Use	33 (9.1%)	25 (6.6%)	0 (0%)	0 (0%)	2 (1.4%)	4 (2.9%)
Current HRT Use	78 (21.4%)	71 (18.8%)	0 (0%)	0 (0%)	-6.50%	8 (5.8%)

Table 5b. Breast Panel Validation Set Patient Characteristics part 2

	In-Situ Cases		Invasive Cases	
Development (D) or Validation (V) Set	D	V	D	V
Stage 1	1 (4.2%)	0 (0%)	67 (48.2%)	51 (36.7%)
Stage 2	1 (4.2%)	0 (0%)	43 (30.9%)	55 (39.6%)
Stage 3	0 (0%)	0 (0%)	18 (12.9%)	13 (9.4%)
Tested for ER	21 (87.5%)	20 (100%)	131 (94.2%)	120 (86.3%)
Tested for PR	21 (87.5%)	20 (100%)	131 (94.2%)	120 (86.3%)
Tested for HER2	1 (4.2%)	0 (0%)	117 (84.2%)	116 (83.5%)
Triple Negative or HER2 Positive	3 (12.5%)	7 (35%)	21 (15.1%)	32 (23%)

A total of seven markers were tested in the entire Breast Panel Validation Set. Each marker was standardized by centering and scaling the results so the healthy controls had a mean of 0 and standard deviation of 1. Linear models were fit to each individual marker in the PDS to evaluate the effects of covariates on serum marker levels. As each marker was standardized prior to fitting the model, coefficients from each model can be interpreted as the expected change in marker level for each unit change of the covariate in units of standard deviations in the healthy controls. For example, HE4 is expected to elevate by an amount equivalent to 0.02 standard deviations in the healthy controls when age is increased by 1 year (table 6a). Results from these models (Tables 6a-b) show how marker levels vary by characteristics of the woman, in particular the effects of malignancy controlling for potentially confounding conditions. Of particular interest is the coefficient for early-stage disease, which is statistically significant only for SPARC. MIC1 also appears to provide some signal in *in situ* disease. However, none of the markers achieved 20% sensitivity at 90% specificity, or performed well in ROC analysis (Table 7).

Age at the time of collection was also found to influence five of seven markers (HE4, FN1, MMP7, COL1A1, and MIC1) and current oral contraceptive (OC) use was found to influence four of seven markers (HE4, FN1, TFF3, and MIC1, tables 6a-b). Hormone replacement therapy and blood draw conditions were found to influence SPARC (table 6b). These factors may have the potential to confound biomarker discovery and validation experiments and should be accounted for in future experiments.

Unpublished Data

Table 6a. Coefficients and p-Values from Linear Models (by GEE) in Breast Panel Development Set (n = 529) for HE4, FN1 and TFF3

	HE4		FN1		TFF3		MMP7	
	Coeff	p-value	Coeff	p-value	Coeff	p-value	Coeff	p-value
Age at Collection	0.0200	0.0020	0.0104	0.0490	0.0092	0.3072	0.0361	0.0000
Current OC Use	0.4870	0.0250	-0.3633	0.0138	1.3800	0.0002	-0.0767	0.5543
Current HRT Use	-0.0239	0.8674	0.0380	0.8052	-0.0320	0.6822	-0.0450	0.8028
Collected Day of Surgery	0.1530	0.4009	-0.1545	0.2703	-0.0508	0.4651	0.0575	0.6387
In Situ	-0.1785	0.2584	0.1299	0.5763	-0.0583	0.6720	-0.2575	0.1464
Early Stage Invasive	0.3596	0.1135	-0.0027	0.9871	0.0837	0.5574	0.1743	0.3405
Late - Early Stage Invasive	0.6440	0.3692	0.1471	0.6019	-0.1216	0.5412	0.1133	0.6737
Triple Negative or HER2 Positive	-0.1928	0.4921	0.3181	0.4031	0.3482	0.2933	0.2364	0.3407

Table 6b. Coefficients and p-Values from Linear Models (by GEE) in Breast Panel Development Set (n = 529) for COL1A1, MIC1, and SPARC

	COL1A1		MIC1		SPARC	
	Estimate	p-Value	Estimate	p-Value	Estimate	p-Value
Age at Collection	0.0220	0.0000	0.0352	0.0000	-0.0089	0.0932
Current OC Use	0.0169	0.8987	0.7605	0.0316	0.0926	0.7087
Current HRT Use	0.2565	0.0955	-0.0702	0.5789	0.3360	0.0354
Collected Day of Surgery	-0.0486	0.6574	-0.0868	0.6287	-0.4297	0.0006
In Situ	-0.3221	0.1989	0.6447	0.0378	0.3746	0.0686
Early Stage Invasive	-0.0935	0.4902	0.1333	0.5146	0.4800	0.0044
Late - Early Stage Invasive	-0.3823	0.0484	0.9238	0.2006	0.2260	0.5062
Triple Neg or HER2 Pos	-0.0412	0.8860	0.1884	0.6657	0.1738	0.5051

Table 7. Summary of ROC Curves in the PDS

	Sensitivity at 95% Specificity		Sensitivity at 90% Specificity		AUC	
	Un-adjusted	Adjusted*	Un-adjusted	Adjusted*	Un-adjusted	Adjusted*
MIC1	0.150	0.059	0.195	0.118	0.538	0.530
FN1	0.044	0.066	0.106	0.160	0.531	0.555
SPARC	0.071	0.049	0.168	0.127	0.560	0.570
COL1A1	0.080	0.066	0.142	0.160	0.608	0.601
HE4	0.062	0.113	0.150	0.160	0.569	0.519
TFF3	0.027	0.075	0.062	0.104	0.460	0.520
MMP7	0.133	0.094	0.177	0.198	0.567	0.545

* Adjusted for Age, Surgical Conditions in First Collection w/o Current OC or HRT Use

Since the 2006 CoE workshop in Arlington, Virginia, resources have been devoted to conducting discovery work in both serum and breast tissue, since there are not enough candidate markers currently ready or available for evaluation. This year CoE investigator Dr. Michèl Schummer completed a discovery project using RNA extracted from tissues of CoE cases and controls to look at expression in genes that have been identified in the

Unpublished Data

literature or by our collaborators as potentially involved in the development of breast cancer. All tissue specimens were characterized by histology and underwent RNA extraction. Dr. Schummer then selected a subset of 94 tissues from 64 participants whose histology corresponded to the patient diagnosis to be used in real-time PCR experiments evaluating gene expression. (Table 8)

Table 8. Specimen Set for Real-Time PCR

Histology	Patients	Specimens
Healthy Control (mammaplasty)	20	28
Ipsi- or contralateral Normal	37	38
Invasive Case	23	25
15 patients donated both cancer and ipsi- or contralateral tissue		

In total, Dr. Schummer has examined 134 genes and identified 46 as possible marker candidates for validation in serum. Genes were ranked by their ability to discriminate between invasive cases and healthy mammaplasty controls using a threshold of 3 standard deviations above the mean of the controls if expression in cases is higher than in controls, or the lowest value of the controls in the case of genes with a lower expression in cases. A transcript marker was terminated in <20% of invasive cancer samples that were higher or lower than threshold. The results of this project are described further on page 16, under “Key Research Accomplishments.” Markers that had serum assays readily available have been evaluated in serum by the TOR Laboratory. So far, 41 protein markers identified by this project have been tested in the BDS. There are 32 marker genes for which no serum assays is commercially available but assay development for 8 of these has been started. The top eight candidate markers identified by PCR were tested in the BDS using available ELISA assays: SPARC, GDF15 (MIC1), FN1[6], SPP1[7], COL1A1, CTGF, MMP7, and WFDC2 (HE4).

The CoE specimen repository is being utilized by collaborators for related discovery and early detection work. For example, CoE collaborator Dr. Samir Hanash of the Fred Hutchinson Cancer Research Center is using specimens from our repository for his project titled, “Alliance of Glycobiologists for Detection of Cancer and Cancer Risk” (U01 CA128427.) The study involves implementation of a new paradigm in the use of glycan biomarkers for early detection of cancer. Dr. Hanash’s Lab used CoE and other specimens for biomarker discovery and contributed 13 candidate markers for evaluation in the BDS. Three of those markers were among the top nine performing markers measured in the BDS and went forward to be measured in the PDS and the PVS to validate these initial findings. Unfortunately, two of the markers (CTGF, FN1) did not perform well in the larger serum set. The third marker (IGFBP4) is still being measured and results are pending.

Dr. Tony Blau received CoE tissue specimens to evaluate the expression of the Epo receptor in breast tumors. His preliminary work helped him to obtain Avon-NCI partnership funding through the SPORE supplemental mechanism. He went on to use CoE tissue samples for laser capture fractionation with the goal of determining the types of cells within primary tumors that express erythropoietin receptor. In addition, he evaluated the presence of a single nucleotide polymorphism (SNP) 1120 bp upstream from the Epo promoter region that has been associated with elevated Epo levels in vitreous fluid. Serum Epo and EpoR levels were measured for the same participants for correlative studies. He has published one manuscript to date[5] and a second one has been submitted. Table 9 below enumerates all collaborators who have received COE specimens and the current outcomes of that work.

Unpublished Data

Table 9. COE Specimens provided to Investigators for Biomarker Measurement

Investigator Requesting Specimens or Information	Specimens Received	Markers Measured	Outcome
Victor Levenson, MD, PhD	BDS, Plasma	Methylation markers	in progress
Samir Hanash, MD, PhD (FHCRC)	BDS, Plasma and Serum	CTGF, CXCL1, FN1, IGFBP2, IGFBP4, IGFBP5	3 markers moved on for measurement in PDS and PVS
Eleftherios Diamandis (Mt. Sinai, Toronto)	BDS, Serum	Two unknown candidates	Not promising
C. Anthony Blau, MD (UW)	Snap frozen and OCT-embedded frozen tissues, FFPE tissues	Epo, EpoR, Epo SNPs	Avon funding, manuscript published
diaDexus, Inc.	BDS	X065 (MIC-1)	Promising - moved on for measurement in PDS and PVS
Andre Baron, MD, PhD (University of Kentucky)	BDS, Serum	sEGFR	in progress
BioCurex	BDS, Serum	RECAF	in progress
Samir Hanash, MD, PhD (FHCRC)	PDS, PVS	TTF3	Not promising

Our consumer advocates have played an important role throughout our CoE. Their activities have included review of new participant materials, participation in scientific meetings, and working with investigators to determine how best a panel of markers could be used in a clinical setting.

During the last funding period we have held four scientific meetings. Recent topics include preliminary results from Dr. Schummer's discovery profiling work, and a presentation by Dr. Melanie Palomares from City of Hope in Duartes, CA on the detection of circulating tumor cells using quantitative real-time PCR. The 2009 CoE Annual Workshop was held in August and the agenda is included as Appendix C. Over the years these meetings have included presentations by study investigators, collaborators and outside experts on work relevant to the aims of the CoE with additional time allowed for discussion. Presentations in 2009 have focused on the results related to study aims 1 and 3.

Investigators continue to refine the draft manuscript of the results from an investigation of the impact of DCIS detection and treatment on breast cancer mortality and associated over diagnosis using a micro simulation model (manuscript is titled Quantifying Risks of Breast Cancer Mortality and Overdiagnosis due to Mammography-diagnosed DCIS.) This is a continuation of work that was initially developed through a previously funded DOD grant (DAMD17-94-J-4237). The primary focus during this final year of the CoE has been on assay development and biomarker evaluation. Investigators plan to finish the micro simulation model manuscript in 2010.

Below we outline each task included in our Statement of Work and detail efforts toward completion of each task. In October 2008, we applied for a no cost-extension extending this study for 12 additional months; therefore, this report represents our entire progress to date for all 7 years of funding (months 1-84).

TASK 1: Recruit women undergoing mammography to donate serial blood samples (Mammography Cohort) (completed)

Task 1a: Obtain Consent to Contact and Screening Questionnaire from women undergoing mammography at participating facilities (complete). This task was

Unpublished Data

conducted between May, 2004 and September 30, 2008. Recruitment of potential participants occurred at Swedish Medical Center mammography clinics and community health and outreach events such as the Swedish SummeRun, which raises money and awareness for ovarian cancer research. During the study period consent to contact was obtained from 1,986 women. These women were sent a short, one page screening questionnaire to collect preliminary risk information. 66% of the screening questionnaires have been returned. Information from these forms was entered into the SIM database and was used to select eligible women to approach for the CoE study.

Task 1b: Obtain mammography data from participating facilities (complete). Collection of mammography data for this study has been completed. Over the course of the CoE we have obtained 5 data downloads from Swedish Medical Center's Mammography Reporting System (MRS), an electronic database used by Swedish Medical Center radiology facilities. For participants who receive mammograms outside of Swedish Medical Center we use self-reported information from the health status or baseline questionnaire to determine the location and date of the woman's most recent screening mammogram and contact the hospital or clinic directly to request a copy of the report and any subsequent diagnostic reports (if applicable.) The reports are then abstracted by trained study staff into the same data entry screens in SIM that store MRS data.

We run a linking algorithm to match study participants to their mammography results from MRS. Using this data we have been able to incorporate mammography information such as assessment code, density and follow-up recommendations into our risk algorithm. Approximately 80% of our participants have electronic records in the Mammography Reporting System.

Task 1c: Using on-going sampling technique, stratify population by risk (complete). Information collected on our study questionnaires and mammography results are used to stratify our study population by risk; that is, allowing us to characterize a woman as high, elevated or average risk. A woman is determined to be at high risk based on family history, if she is of Ashkenazi Jewish descent, self-reports a positive test for the BRCA 1 or BRCA 2 mutation, or has prior history of receiving a breast biopsy. A woman is determined to be at elevated risk for breast cancer by GAIL Model, breast density, mammography assessment codes, or mammography follow-up recommendations. The table below summaries the number of women enrolled into the mammography cohort and associated risk levels based on collected information. All women reported in the table have donated specimens and completed baseline questionnaires.

Table 10. Mammography Cohort Breakdown by Risk

Mammography Cohort participants donating specimens with completed baseline data		
Risk Category	Participants	Percentage of Total
High	233	37.4%
Elevated	337	54.1%
Average	53	8.5%
Total	**623**	**100%**

Task 1d: Approach selected women for blood donation (complete). Participants were approached for blood donation beginning in October, 2004 and ending in September, 2008. Participants are asked to donate blood on or close to the date of their annual screening mammogram. Of the 623 mammography cohort participants who have donated blood for the study, 449 completed serial draws.

Task 1e: Send blood donation appointment letters and epidemiologic risk factor questionnaires to consenting women (complete). To date, 637 women have completed one or more study blood draws. All of these women received an epidemiologic risk factor

Unpublished Data

questionnaire at baseline and a shorter health status questionnaire at each blood draw appointment (provides updated information about medical history variables that may affect marker status). 100% have completed and returned the health status questionnaire and 601 (96%) completed and returned the baseline questionnaire. Updated family history information is collected on an end-of-study supplemental questionnaire mailed out after the participant's final study blood draw.

Task 1f: Receive and data enter questionnaires (complete). Questionnaires are entered within 48 hours of receipt. Quality control data entry is performed on all baseline questionnaires and approximately 10% of the health status questionnaires. The database manager periodically reviews quality control data entry on all questionnaires to track our error rates and ensure quality control entry is occurring on a sufficient number of questionnaires.

TASK 2: Recruit women undergoing stereotactic biopsy to donate pre-biopsy and serial follow-up blood samples (Biopsy Cohort)

Task 2a: Finalize approach procedures to be used by Swedish Breast Care Center (complete). In September 2001, Dr. Urban received funds from an NCI-Avon "Progress for Patients" award (P5OCA83636) that allowed us to develop and test procedures to recruit and enroll women who were undergoing stereotactic biopsy at the Swedish Breast Care Center (SBCC), part of SMC. For this "Avon study" women were asked to provide a one-time, pre-biopsy blood donation and complete both the baseline and health status questionnaires. 143 women were enrolled in this study at SBCC. We adopted the same procedures to recruit and enroll women scheduled for breast biopsies at SBCC into the CoE study. Women enrolled into the CoE were asked to give a blood sample prior to their biopsy procedure *in addition* to an annual sample at the time of subsequent mammograms.

Task 2b: Specimen Collection Specialist attends biopsy appointment to obtain informed consent, collect pre-biopsy blood sample, and provide epidemiologic risk factor questionnaire (complete). Biopsy patients were approached and enrolled in this study from January, 2007-August, 2008. To date 13 participants have been enrolled and donated a blood sample just prior to their biopsy (1 invasive case, 2 in situ and 10 benign controls).

TASK 3: Recruit women undergoing surgery to donate pre-surgery and follow-up blood samples, and collect tissue on selected breast cancer cases (Surgical cohort).

Task 3a: Work with surgeons' offices to integrate patient approach procedures into the patient care flow. (complete). We have worked closely with participating breast surgeons and clinic staff to design and implement patient approach procedures for recruitment that have proven to be successfully integrated with normal clinic flow. Our study personnel are able to maintain an open dialogue with participating physicians about study progress and procedures by checking in with them and their staff on a daily basis. This creates an environment where physicians and study staff are able to work together to continuously refine and improve our approach procedures.

Task 3b: Pilot patient approach and specimen collection procedures (complete).

Patient approach began in July, 2004. Swedish Medical Center breast surgeons identify patients that are likely candidates for surgical specimen collection and at the pre-surgical visit approach these patients about study participation. If the patient is interested, the physician will obtain verbal consent for study staff to contact the patient either in person or by phone. If a study staff member is present at the clinic, the physician invites the woman to speak to the study representative who can help answer immediate questions or concerns. If the patient chooses, she may be enrolled at this time (if she meets the

Unpublished Data

eligibility requirements). Otherwise, study staff contacts her by phone to discuss the study in further detail and set up an enrollment appointment to conduct in-person informed consent and collect a pre-surgical blood sample.

Task 3c: Routinely approach selected women undergoing surgery for blood and tissue collection or blood only collection (complete). Surgical patients in Seattle were approached and enrolled in the CoE from July, 2004-September, 2009. Since April, 2008 recruitment has been limited to only those patients whose specimens were still needed for the discovery work described on page 10. To date, we have enrolled 90 participants in Seattle from SMC who have successfully completed questionnaire data and donated blood and tissue, and 296 who have completed questionnaire data and donated only blood.

Task 4. Recruit women undergoing biopsy or surgery to donate a one-time only pre-surgical blood *and* tissue sample, as feasible, at Cedars Sinai Medical Center.
Task 4a: Finalize approach procedures to be used by Dr. Scott Karlan at Cedars-Sinai Medical Center (complete). This task has been completed and the Cedars-Sinai Clinical and Recruitment protocol received DoD Human Subjects approval in July 2005.
Drs. Scott and Beth Karlan have approached physicians who attend Breast Center conferences, to educate them about available research protocols for interested patients. Recruitment flyers and brochures are posted around the Cedars Sinai campus (specifically, the Saul and Joyce Brandman Breast Center and the Cedars-Sinai Outpatient Surgery Center) and made available to raise patient awareness. This study is also listed on the Cedars-Sinai web site.

Eligible women previously scheduled for a breast surgical procedure that involves the removal of some or all of their breast tissue are approached about possible study participation. Patients are not scheduled for surgical procedures for the purpose of this study alone. The Principal Investigator, co-investigators, or treating physicians (usually a breast surgeon, occasionaly a radiologist or a medical oncologist) help identify potential subjects. The treating physician makes initial contact with potential subjects and contacts a trained study staff member to consent the patient into the study if the woman agrees to participate.

Task 4b: Routinely approach selected women for blood and tissue collection (complete).
Drs. Beth and Scott Karlan and their study staff recruited eligible women into the COE study at Cedars Sinai Medical Center from October 2005-April, 2008. Their study enrollment goal was 50 surgical women per year for the duration of the study. The population includes healthy women with no disease, women with benign lesions and pre-malignant breast diseases, and women with in-situ and invasive carcinoma. Of the 77 surgical participants who completed the baseline questionnaire, 69 donated both blood and tissue and 8 donated only blood.

Task 4c: Surgeon to collect healthy tissue, benign lesions, atypia, in situ disease, and invasive carcinoma tissue samples. (Complete). The Cedars Sinai team has implemented the shared tissue collection protocol and has collected tissue samples from 196 study participants. Pathology information is centrally abstracted at FHCRC using a Patient Level Clinical Diagnosis form.

Immediately after the surgeon has removed the necessary tissue and the pathologist has taken what is required for pathologic diagnosis, a study Specimen Collection Specialist is permitted to collect specimens from the removed tissue for the purposes of the CoE. All or part of the un-needed tissue is collected, labeled and processed for storage. The tissue is embedded in OCT and/or snap frozen. Tissue collected includes malignant tissue with adjacent normal tissue, as well as tissue from pre-malignant

Unpublished Data

lesions and breast tissue from normal patients undergoing plastic surgery procedures at Cedars-Sinai.

The Patient Level Clinical Diagnosis form uses information that has been abstracted from pathology and other medical reports to characterize a woman based on TNM staging and grade of disease at the time of her diagnosis. A FHCRC study staff member completes this form for all CoE surgical participants with the research nurse conducting quality assurance.

Breast Tissue Histology Review

Working closely with breast pathologists Drs Sean Thornton and Ellen Pizer of Washington Pathology Consultants, we have created two forms: Breast Histology Tissue Review Form and a Clinical Status Follow-Up Form that are used to characterize breast tissue samples and capture treatment and disease status post-diagnosis and surgery. In 2006 a review of a pilot group of 60 tissues from 10 participants was conducted in an effort to discover whether the most severe diagnosis listed on the patient pathology report matched the actual histology of the tissue specimens collected. The results (reported in the 2007 annual progress report) led us to conclude that tissue specimens must be independently reviewed to accurately determine histology; it is insufficient to rely solely on the pathology report from surgery. At this time, review is only performed on tissues being used for discovery.

Task 5. Blood samples from Mammography, Biopsy and Surgical Cohorts are collected, processed into serum and plasma cryovials, and logged into specimen tracking system (complete).

In all blood collections, the Specimen Collection Specialist collects up to 50 ml of whole blood. At the initial collection the phlebotomist will distribute the blood between 3 red top (serum) tubes, 1 purple top (EDTA plasma) tube, and one yellow top (ACD-plasma and lymphocytes) tube. For all subsequent draws, blood is collected in 4 red top tubes and 1 purple top tube.

Standard protocols are followed to process specimens into sera and plasma and aliquot them into cryovials uniquely labeled with study specimen ids. Specimens are then logged into the Specimen Tracking System database (STS).

The blood specimens are stored in 1 ml quantities to avoid damaging freeze-thaw cycles. Aliquoted specimens are entered into the specimen tracking system then transported to the study repository for long-term storage and will eventually be delivered to laboratory investigators for future analysis. Blood draw date and time, and time of processing and freezing are recorded in STS as well.

Task 6. Revise existing ovarian cancer database to accommodate breast tissue specimens and questionnaire

Task 6a: Analyze **current system** and prepare preliminary assessment of revised software design specifications (complete). FHCRC programmers have enhanced an existing specimen tracking system (STS) to accommodate specimens and breast specimen data being collected as part of the CoE. We currently track the following specimen data: date of blood and/or tissue donation, specimen processing, amount of specimen collected, types of specimen storage, and storage location of specimen aliquot or tissue vial or block.

Task 7. Develop an implementation test utilizing proposed software with a middle tier and internet interface for the Clinical Data Module (complete).

Infrastructure in place includes: web server hardware, web service software, access security, data entry form templates, and referential integrity between database objects.

Unpublished Data

We have refined an Access database to track information that is collected on our Patient Level Clinical Diagnosis Form. This form provides appropriate information to characterize a woman based on TNM staging and grade of disease at the time of her diagnosis. It also captures receptor status information, such as estrogen or progesterone positivity/negativity, which will be used to select specimens for the different specimen sets. This database acts as our "clinical module" and is linked to SIM, our primary data management system, which in turn is linked to the STS and SpecimenDB (see task 8 below for more detail.)

Web based screens in SIM for questionnaire data entry and patient tracking have been developed. Routines for data validation with each submission of data to the server have been implemented. Every value entered is checked for validity. Any outliers are returned to the data entry specialist for verification before the data are committed to the database. In addition, attempts to re-enter data that have previously been collected, are preempted via referential integrity.

Task 8. Develop breast specimen tracking database to replicate and enhance the current system's functionality adjusting per information gained in the implementation test (complete).

In 2006 Staff Scientist Dr. Michèl Schummer developed SpecimenDB, a FileMaker database for information that is generated from our specimens, such as experimental and specimen processing results. SpecimenDB also serves as a front-end to CoE databases SIM, STS and the Access database tracking our Patient Level Clinical Diagnosis Form. The interface provides a unified look across all components and is thus easy to navigate. Each field can be searched without knowledge of the underlying structure. Summary reports can be generated from any view as Excel or PDF documents. SpecimenDB is client- and web-based, the latter allowing for collaboration across sites. Although the back-end consists of several databases, the user sees just three major areas: Specimens, Patients and Results.

The Specimens area holds data about the processing of the specimens, such as RNA extraction (Figure 3a.) This allows for technicians to enter information pertaining to specimen processing. Having this information in a central location will prevent us from distributing a specimen that was previously known to yield poor RNA or protein. The Specimens area also has a view that lists multiple specimens in rows which allows for intuitive searches and the generation of summary reports.

The Patients area holds patient-related information that has been stripped of identifying information, including the pathology reports, both abstracted and a scanned copy. Similar to the specimen area, it is possible to toggle between views that list detailed information about a single or multiple patients. In list view, it is further possible to toggle between patients and their specimens, allowing for simultaneous querying of patient and specimen information.

The Report area is designed to contain experimental data obtained from the specimens in our repository. We have designed a database module (written in FoxPro) that keeps track of our serum and plasma marker measurement workflow, including the results. Although optimized to work with our laboratory, this module can also accommodate results data generated in other laboratories. We are currently expanding SpecimenDB capabilities to link to these results. This will allow us to perform queries across patients, specimens and results simultaneously in an extremely user-friendly manner. The new, integrated view will look very similar to the current view.

Task 9. Develop collaborative web site

Unpublished Data

Task 9a: Develop site to support real-time discussion and information sharing among investigators (complete). Investigators and study staff continue to utilize the CPAS website, created by Dr. Martin McIntosh and his Computational Proteomics Laboratory group at the Fred Hutchinson Center. CPAS is an open-source science portal offering web-based bioinformatics and collaboration tools to help scientists store, analyze and share data from high-throughput experiments and clinical trials[8]. CPAS is available as free, installable software, with source code. This work was funded by NCI subcontract 23XS144A. Study investigators and staff use CPAS to support real-time communication and information sharing among FHCRC staff, CoE investigators and their respective staff. A username and password are required to access information on this site. The content on CPAS is organized hierarchically into projects and subfolders, much like the file directories on your computer; therefore, users find it easy to navigate through and use.

In addition to CPAS, from July,2004-July, 2009 we maintained a second website to function as a study reference to outside researchers and the general public. The website consisted of four main sections: a Homepage, Research Overview, Advocacy, and Community Events. There was also a link to our internal CPAS site accessible only to project investigators and staff.

Task 9b: Develop extensions that will give investigators ability to query specimen tracking system and download summary reports (complete). SpecimenDB (explained in detail under Task 8) tracks both specimen and patient related (clinical) information. Its unification of several databases allows investigator-generated queries. For example, a user can select patients that match certain clinical criteria and click on the "toggle specimens" button. Available specimens matching the criteria will be shown. The user can then search for subsets of these specimens, such as available serum volume. Queries can be performed in increments, which will allow the investigator to review the data between steps. Multiple AND or OR statements can be applied without knowledge of the underlying database structure. Once a subset of records has been identified, a summary report can be generated through pre-configured templates, or ad-hoc, through user-selection. To facilitate this process, field names are the same in the user interface as in the underlying database. In addition, the CoE CPAS site is linked to the study's data management system; therefore, investigators are able to access and view data reports as if they were in the SIM system.

Task 9c: Develop web pages for each investigator that are linked to collaborative site (complete). We have developed folders on CPAS for each laboratory based investigator. Each investigator is able to design their own folder and create subfolders suiting their specific needs; however, we request that investigators use their folders to upload all laboratory results and to view marker results. We have also developed folders to support investigator specific meetings and collaborative activities, such as the quarterly investigator calls and the developing Specimen Review Committee. In addition, we have created a folder that is open to the public to support the upcoming CoE investigator meetings.

Task 10: Prepare and Analyze BDS

Task 10a: Provide samples from a set of 66 women to laboratory investigators (complete). BDS specimens were provided to study laboratory investigators to determine the preliminary usefulness of new markers. Characteristics of the participants whose specimens make up the BDS are described above in table 4. The assays used to measure markers in the BDS are described in Appendix A.

We will apply for future funding to continue working with our collaborators listed in Appendix B to identify new markers and test their efficacy in the BDS. New funding will

Unpublished Data

also allow promising markers to continue through the pipeline and be evaluated in the PDS and PVS. In addition the BDS and the Breast Panel Validation Sets are available to new collaborators that may have promising markers or discovery platforms through an RFA mechanism.

Tasks 10b and 10c: Statistical analysis of preliminary BDS results (complete). Statistical analyses on preliminary assay results from the BDS were completed this year and the results are summarized on page 7 in Figures 1 and 2.

Tasks 11-13: Biomarker Panel Validation and Evaluation (months 55-84)

Task 11. Prepare and analyze Panel Development Set Markers were measured in both the PDS and the PVS simultaneously. Statistical analyses have been performed only on the PDS results and that is what is reported below for tasks 12 and 13.

Task 12. Conduct statistical work to evaluate candidate biomarkers (in PDS)
Task 12a. Establish cut offs for normals (complete). The following markers were evaluated: COL1A1, FN1, HE4, MIC1, MMP7, SPARC and TFF3. Thresholds were estimated using the PDS for 90% and 95% Specificity levels.

Task 12b. Assess single marker sensitivity and specificity for candidate biomarkers (complete). ROC analyses were conducted to evaluate sensitivity and specificity of individual markers in the PDS.

Task 12c. Examine stability of markers over time within and between subjects. Effects of covariates on marker levels as well as the within and between woman variance of markers were estimated. Generalized estimating equations (GEE) methods were used to account for correlation of results from multiple blood collections within the same women at separate time points. Stability over time and effects of covariates on the marker levels were evaluated using GEE and also by graphical analyses.

Task 12d and 12e. Using augmented logisitic regression, estimate optimal combinations of markers in a longitudinal setting and use an ROC curve to evaluate the contribution of markers to mammography. We are unable to complete this task because individual markers did not have sufficient sensitivity or specificity in the PDS to warrant either investigating the benefit of combining markers in a panel or their ability to complement mammography, therefore this task will not be completed.

Task 12f. Provide feedback to laboratory scientists via CPAS each step of the way
While CPAS was useful in other areas of this study, it was found that email and in-person meetings were a more effective tool for communicating feedback to our lab scientists.

Task 13. Prepare and analyze Panel Development and Validation Sets (PDS, PVS)
Task 13a. FHCRC Laboratory technician to conduct biomarker assays on blinded samples from 500 women in the PDS (complete).Seven markers (listed above) were measured in both the PDS and the PVS simultaneously.

Task 13b. Blinded samples given to laboratory scientists to continue refinement of new assays. Due to the heterogeneity of breast cancer subtypes, preliminary analyses were conducted in the PDS which was comprised of specimens from all cases meeting certain data availability requirements regardless of subtype. Laboratory scientists remained blinded to the PDS and PVS at all times, so this set was un-useful for optimizing assay conditions. To focus laboratory efforts on more homogeneous groups of specimens, we

Unpublished Data

are identifying subgroups of CoE specimens that may be relevant to specific biomarker discovery experiments or assay optimization using other sources of funding.

Tasks 13c. and 13d. Data to Dr. McIntosh to validate the ability of the marker panel to discriminate breast cancer from non-cancerous conditions and biomarker validation team to evaluate the improvement in performance attributable to marker panel. Individual markers did not have sufficient sensitivity or specificity in the PDS to warrant investigating the benefit of combining markers in a panel.

Task 13e. Prepare reports and manuscripts describing performance of marker panel. In progress. Manuscripts are currently being drafted describing the PDS and PVS specimen sets, methods for assays used in analyzing these sets, and findings from the PDS. We have one completed manuscript (Appendix D) described in more detail below under key research accomplishments.

Key Research Accomplishments

- Assays for 83 potential biomarkers were purchased or developed and tested in the BDS by the TOR laboratory and collaborators.
- Assays for seven biomarkers were tested in the Panel Development and Validation Sets by the TOR laboratory.
- Due to a lack of candidate markers available for evaluation we have devoted project resources to discovery in CoE tissue samples. We conducted a comparison of gene expression profiles in tissues collected from surgical cohort cases and from healthy controls undergoing reduction mammaplasty. Potential molecular targets for differential expression were identified by a) mining publicly available expression data and b) utilizing a commercial PCR array. 46 genes with differential expression between cases and controls were identified. In cases, 7 of 38 normal tissues removed from a distant site in the diseased breast exhibited a cancer-like expression profile. The remaining 31 tissues were genetically similar to the profiles from samples collected from mammaplasty controls. This suggests it may be possible to identify regions of ipsilateral histologically “normal” breast tissue that are predisposed to malignancy. These areas could then be targets for localized treatment for prevention. Most importantly, 12 genes were discovered with under-expression in cancers linked to aggressive disease with poor outcomes. These genes were not previously associated with breast cancer and have the potential to become markers of prognosis. These results are described in a new manuscript (Appendix D) submitted September 23, 2009 to the Journal American Association of Cancer Research titled, “The Discovery of Novel Human Breast Cancer Markers with Potential for Prognosis in Early Detection.”

Reportable Outcomes

October 2008-September 2009

1) Miller CP, Lowe KA, Valliant-Saunders K, Kaiser JF, Mattern D, Urban N, Henke M, Blau CA. Evaluating Erythropoietin-Associated Tumor Progression Using Archival Tissues from a Phase III Clinical Trial. Stem Cells. 2009 Sep;27(9):2353-61. (Appendix D)

Previously Reported Outcomes (October 2002-September 2008)

1) Scholler N, Garvik B, Quarles T, Jiang S, Urban N. Method for generation of in vivo biotinylated recombinant antibodies by yeast mating. J Immunol Methods. 2006 Dec 20;317(1-2): 132-43.

Unpublished Data

2) Urban ND, Longton GM, Crowe AD, Drucker MJ, Lehman CD, Peacock S, Lowe KA, Zeliadt SB, Gaul MA. Computer-Assisted Mammography Feedback Program (CAMFP): An Electronic Tool for Continuing Medical Education. *Academic Radiology*. 2007 Sep;14(9):1036-42

3) Loch, C. M., Ramirez, A. B., Liu, Y., Sather, C. L., Delrow, J.J., Garvik, B., Scholler, N., Urban, N., McIntosh, M. W. and Lampe, P. D. Use of High Density Antibody Arrays to Validate and Discover Cancer Serum Biomarkers. *Molecular Oncology.* December 2007. Vol. 1, Issue 3, Pages 313-320.

4) Thorpe, JD, Duan X, Forrest R, Lowe K, Brown L, Segal E, Nelson B, Anderson G, McIntosh M, Urban N. Effects of Blood Collection Conditions on Ovarian Cancer Serum Markers (2007) *PLoS One* 2(12): et1281.doi:10.1371/journal.pone.0001281.

5) Use of cancer-specific yeast-secreted *in vivo* biotinylated recombinant antibodies for serum biomarker discovery; Scholler, N, Gross, JA, Garvik, B, Wells, L, Liu, Y, Loch, CM, Ramirez, AB, McIntosh, MW, Lampe, PD, Urban, N. Journal of Translational Medicine 2008, 6:41.

6) Use of yeast-secreted in vivo biotinylated recombinant antibodies (biobodies) in bead-based ELISA; Scholler N, Lowe K, Bergan L, Kampani A, Ng V, Forrest R, Thorpe J, Gross J, Garvik B, Drapkin R, Urban N. Clinical Cancer Research, 14 (9): 2647. (2008).

Conclusions

We have quite exhaustively searched without success for an early detection serum marker for breast cancer. We have learned that markers that initially appear promising are shown in a careful validation study to be influenced more by confounding conditions than by malignancy. Five of 7 best candidate markers are influenced by age of the woman (MIC1, COL1A1, HE4, FN1 and MMP7). Similarly, 4 of 7 markers are affected by current OC use (TFF3, FN1, HE4 and MIC1). One of the markers (SPARC) is affected by current HRT use as well as conditions of the blood draw (surgical vs. clinic visit). We conclude that future discovery efforts must account for these confounding factors to avoid identification of markers for hormone use or age rather than malignancy. The well-annotated specimens that we have collected will be useful for such discovery efforts as well as for further validation efforts. Half of our validation set remains blinded so that it can be used when markers are eventually identified that are worthy of inclusion in a panel.

References

1) Gail Model Risk Factors: Impact of Adding an Extended Family History for Breast Cancer; Anna Crispo, Giuseppe D'Aiuto, MariaRosaria De Marco, Massimo Rinaldo, Maria Grimaldi, Immacolata Capasso, Alfonso Amore, Cristina Bosetti, Carlo La Vecchia, and Maurizio Montella. The Breast Journal. Volume 14 Number 3, 2008; 221-227.

2) Mammographic Density Correlation with Gail Model Breast Cancer Risk Estimates and Component Risk Factors. Melanie R. Palomares, Joelle R.B. Machia, Constance D. Lehman, Janet R. Daling and Anne McTiernan. Cancer Epidemiol Biomarkers Prev 2006; 15(7). July 2006.

Unpublished Data

3) Increased levels of SPARC (osteonectin) in human breast cancer tissues and its association with clinical outcomes; Gareth Watkins, Anthony Douglas-Jones, Richard Bryce, Robert E Mansel, Wen G Jiang. Prostaglandins Leukot Essent Fatty Acids. 2005 AprL 72(4): 267-72.

4) Extracellular maxtrix signature identifies breast cancer subgroups with different clinical outcome. Bergamaschi A, Tagliabue E, Sørlie T, Naume B, Trulzi T, Orlandi R, Russnes HG, Nesland JM, Tammi R, Auvinen P, Kosma VM, Ménard S, and Børresen-Dale AL. J Pathol. 2008 Feb; 214(3): 357-67.

5) Miller CP, Lowe KA, Valliant-Saunders K, Kaiser JF, Mattern D, Urban N, Henke M, Blau CA. Evaluating Erythropoietin-Associated Tumor Progression Using Archival Tissues from a Phase III Clinical Trial. Stem Cells. 2009 Sep;27(9):2353-61.

6) Fibronectin Expression Modulates Mammary Epithelial Cell Proliferation during Acinar Differentiation. Courtney M. Williams, Adam J. Engler, R. Daniel Slone, Leontine L. Galante and Jean E. Schwarzbauer. Cancer Res 2008; 68: (9). May 1, 2008.

7) Osteopontin overexpression in breast cancer: knowledge gained and possible implications for clinical management. Tuck AB, Chambers AF, Allan AL. J Cell Biochem. 2007 Nov 1; 102(4): 859-68.

8) Installation and use of the Computational Proteomics Analysis System (CPAS). Myers T, Law W, Eng JK, McIntosh M. Curr Protoc Bioinformatics. 2007 Jun; Chapter 13: Unit 13.5.

Appendices

Appendix A	Summary of CoE assay development work
Appendix B	Agenda for the 2009 CoE Workshop
Appendix C	submitted manuscript: The Discovery of Novel Human Breast Cancer Markers with Potential for Prognosis in Early Detection
Appendix D	publication: Evaluating Erythropoietin-Associated Tumor Progression Using Archival Tissues from a Phase III Clinical Trial. Stem Cells.
Appendix E	List of personnel receiving pay

Unpublished Data

In-house assays

Assay	Selected By	Developed By	Plate-based or Bead-based?	Capture Source	Detection Source
COL1A1 (Collagen I)	*M. Schummer*	LB (TOR)	Bead-based	ABCAM	Santa Cruz
***CTGF**	*S. Hanash*	AK (TOR)	Bead-based	Antigenix	Antigenix
HE4 (WFDC2)	*M. Schummer*	AK (TOR)	Bead-based	Covance	In-House
MIC1 (X065)	*DiaDexus*	LB (TOR)	Bead-based	DiaDexus	DiaDexus
SPARC (Osteonectin)	*S. Sukumar*	LB (TOR)	Bead-based	R&D Systems	GeneTex
****TTF3**	*S. Hanash*	**See reference	Plate-based	Abnova	R&D Systems

* Antigenix development kit was used to develop/optimize an in-house bead-based assay.

** Screened by SP (Hanash). Protocol adapted from: Bignotti E, et al. Trefoil factor 3: a novel serum marker identified by gene expression profiling in high-grade endometrial carcinomas. Br J Cancer. 2008 September 2; 99(5): 768–773.

Commercial assays - performed according to manufacturers instructions

Assay	Selected By	Plate-based or Bead-based?	Kit Source
IL-17	Part of a cytokine panel	Bead-based	LINCO
****MMP-7 (used in BDS)**	*M. Schummer*	Plate-based	R&D Systems
****MMP-7 (used in BPVS)**	*M. Schummer*	Bead-based	R&D Systems
Osteopontin (SPP1)	*P. Nelson*	Plate-based	R&D Systems

*** The plate-based assay was compared to the bead-based assay in the OTS. The two assays correlated very well. Therefore, the bead-based assay was used for the BPVS.

Commercial assays - additional optimization required

Assay	Selected By	Plate-based or Bead-based?	Kit Source
FN1 (Fibronectin 1)	*S. Hanash*	Plate-based	Bender MedSystems
IGFBP-4	*S. Hanash*	Plate-based	R&D Systems

Unpublished Data

CENTER OF EXCELLENCE ALL-INVESTIGATOR WORKSHOP

August 7, 2009

Center for the Evaluation of Biomarkers for Early Detection of Breast Cancer
Fred Hutchinson Cancer Research Center, Seattle, WA
Day Campus, Arnold Building 1st Floor M1- A303/305/307 *Questions? Please call 206-667-4238*

8:00 am	**CONTINENTAL BREAKFAST**
8:30 am	**Opening Remarks** Nicole Urban
8:45 am	**Biomarker Discovery** Michèl Schummer, Scott Karlan, Beth Karlan
9:45 am	**Assay development and biomarker evaluation in serum** Robin Forrest, Archana Kampani, Lindsay Bergan, Jason Thorpe
10:45am	**BREAK**
11:00 am	**Evaluation of top candidates in the Panel Validation Set** Nicole Urban, Martin McIntosh, David Beatty
12:00 pm	**LUNCH**
1:00 pm	**Discussion regarding organization of material for papers**
2:15 pm	**Discussion regarding future directions for this work**
3:30 pm	**CLOSING**

TITLE

The discovery of novel human breast cancer markers with potential for prognosis and early detection

AUTHORS

Michèl Schummer [1], Ann Green [1], J. David Beatty [2], Beth Y. Karlan [3], Scott Karlan [4], Jenny Gross [3], Sean Thornton [5], Martin McIntosh [1], Nicole Urban [1]

[1] Molecular Diagnostics Program, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

[2] Swedish Cancer Institute, 1221 Madison Street #400, Seattle, WA, 98104, USA

[3] Women's Cancer Research Institute at the Samuel Oschin Comprehensive Cancer Institute, Cedars Sinai Medical Center, Los Angeles, CA 90048, USA

[4] Saul and Joyce Brandman Breast Center, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA

[5] Cellnetix, 1124 Columbia Street, Suite 200, Seattle, WA 98104, USA

RUNNING TITLE

Breast cancer marker discovery for prognosis and detection

Unpublished Data

KEY WORDS

Breast cancer, PCR, breast reduction mammaplasty, prognosis, early detection

NOTES

This work was supported by DOD/CDMRP DAMD17-02-1-0691 and by NCI/Avon P50 CA083636 (all authors).

Requests for reprints: Michèl Schummer, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, M2-B230, Seattle, WA 98109, USA

Unpublished Data

ABSTRACT

Mammography reduces breast cancer mortality in women over 50, but its performance is often criticized. Finding a marker (or markers) to complement mammography, thus improving its sensitivity and specificity, would have great clinical value. Unfortunately, no serum marker has proven reliable in detecting breast cancer. To find such a marker, we utilized two sets of well-characterized tissues: one from breast cancer patients and the other from healthy women undergoing reduction mammaplasty. We identified over 46 differentially expressed genes from a large list of potential targets by a) mining publicly available expression data (identifying 134 genes for quantitative PCR) and b) utilizing a commercial PCR array. These genes warrant further investigation as potential blood markers for early detection. As a second finding, when histologically normal breast tissue was removed from a distant site in a breast with cancer, specimens from 7 of 38 patients displayed a cancer-like expression profile, while the remaining 31 were genetically similar to the reduction mammaplasty control group. This suggests that it may be possible to identify regions of ipsilateral histologically ‘normal’ breast tissue that are predisposed to becoming malignant. This might lead to future clinical methodologies to identify normal-appearing tissue that warrants localized treatment for prevention. Most importantly, 12 genes showed lower expression in cancers with a poor outcome, suggesting their use as prognostic markers; these genes were also underexpressed in a large number of controls.

Unpublished Data

INTRODUCTION

While mammography reduces breast cancer mortality in women over the age of 50 (1, 2) there is controversy regarding the degree of benefit (3). Most critics agree that mammography has less value in women under 50 (4) due to lower sensitivity and a high rate of false positives (5). Existing serum markers (CA 15-3, CEA and CA 27-29) have both low sensitivity and specificity. Although they may be useful for monitoring treatment in patients with advanced disease (6), these markers are not helpful in the early detection of breast cancer. There is thus a pressing need for novel markers that can be used independently or that can complement mammography in early detection.

In 1999 we successfully used a transcript-based discovery approach to identify early detection markers for ovarian cancer (7). Following the 5 phases of screening biomarker development proposed by Pepe *et al.* (8), HE4, the product of the human epididymis gene WFDC2, was developed into a serum assay (9, 10) that is now approved for remission monitoring of ovarian cancer. It is being evaluated for its potential role in screening and is considered a successful product of translational biomarker research.

Because breast cancer marker research has focused mainly on prognosis (6), there are few comprehensive studies to identify early detection markers. We therefore relied on our previously successful approach using gene discovery by cDNA microarray followed by expression validation through polymerase chain reaction (PCR), ranking of potential markers and the development and testing of serum assays. For breast cancer, a large body of research already available in the public domain allowed us to forego our

Unpublished Data

own microarray work; instead we mined publicly available expression data in tissues of breast cancer and normal healthy controls.

One of the lessons learned from previous gene discovery experiments is the importance of having high-quality appropriately preserved specimens and matching patient data. We therefore spent considerable effort on the accrual of needed tissues. In close collaboration with participating surgeons and pathologists, we were able to collect specimens in the operating and gross rooms where they were processed with as little delay as possible, thus minimizing variability. In addition, routine clinical gross and microscopic tissue analysis was complemented with routine research histological examination on the actual tissue piece that was later used for expression analysis. Breast tissues from breast cancer patients were then compared to those from healthy individuals. While normal tissue adjacent to the cancer is relatively easy to obtain, we feared that in cancer patients, cancer-related pathways may be perturbed in these tissues (11). We therefore used normal tissue from breast reduction mammaplasties as controls.

The identified genes have the potential to become markers for molecular pathology (e.g. aiding the pathologist deciding about the malignant potential of a suspicious-looking tissue section), prognosis, guiding therapy and for early diagnosis (as proteins in serum), in a panel, potentially complementing mammography.

Unpublished Data

METHODS

Patients and tissues. Patients were enrolled at Swedish Medical Center, Seattle and Cedars Sinai Medical Center, Los Angeles. Patients were consented before surgery and administered a health status and family history questionnaire. Hospital records were used for follow-up. Cancer, ipsi- and contralateral normal tissues were obtained at Swedish Medical Center from 44 patients (including 7 with neoadjuvant treatment) undergoing mastectomy. Tissues from breast reduction surgeries (20 patients) were obtained from private practices in Los Angeles and histologically analyzed to exclude any with abnormalities. In all cases, tissue was obtained and processed by research personnel in the operating or gross room and frozen within 1 hour of surgery. The frozen tissue available for research (mean: 150 g, range: 20-500 g) was split into several pieces of which one was fixed in formalin, embedded in paraffin and used for histological examination by a pathologist. The other pieces were kept frozen and used for RNA extraction. Only tumor samples with more than 70% tumor cells, excluding *in situ* disease, and normal samples with less than 60% fat were included. In the end, gross and microscopic clinical evaluation matched the histology of the actual tissue piece being analyzed in 50% of the cancer tissues and 67% of tissues with normal histology. Patient characteristics are reported in the supplemental Table S1.

LevelsDB. Over the last 10 years a database has been compiled (LevelsDB) that holds gene and protein expression information from over 134 publications (90% transcript-, 10% protein-based) and 21,890 genes. LevelsDB was created to facilitate the discovery

Unpublished Data

of markers for cancer detection, and emphasis was given to publications with data for normal controls as well as cancers. LevelsDB uses the GeneID as an identifier (12) which is related to the gene symbols governed by the Guidelines for Human Gene Nomenclature (13). The datasets were extremely variable in the way they recorded expression, ranging from a simple list of proteins to raw cDNA microarray expression data. As a consequence, LevelsDB forewent exact representation of original expression values. Instead, it recorded whether or not a transcript or protein was present in a given tissue, whether it had tumor-to-normal ratios above a factor 2 or, in the case of cDNA microarray-based expression data, whether a mRNA was expressed at low, medium or high levels (threshold defined by 1x, 3x, and 10x the median expression across all tissues). LevelsDB also contains data on subcellular localization. The datasets used in LevelsDB are listed in the Supplement.

RNA extraction and real-time PCR (SYBR). Snap-frozen tissues were homogenized with a TissueLyser (Qiagen, Valencia, CA) in Trizol (Invitrogen, Carlsbad, CA). Total RNA was then extracted using RNeasy with DNAse I (Qiagen). RNA quality was measured by Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA) to have a 28S/18S RNA ratio of 1 ± 0.2 and by spectrophotometer with an OD_{260}/OD_{280} ratio > 1.6). Mean RNA yield was 90 ± 130 ng per mg of tissue. Copy DNA was reverse transcribed from 5 µg of total RNA (Superscript III kit, Invitrogen) with oligo-dT priming, of which 50 ng were used as template in a 15 µl PCR. Copy DNA was amplified using the SYBR green kit (Invitrogen) on a 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, CA). Each 384-well plate contained aliquots of all cDNAs used during the experiment as well

Unpublished Data

as a standard made with testes cDNA in 5 dilutions (1:1, 1:3, 1:9, 1:27, 1:81) as duplicates, amplified with primers for ACTB (all cDNAs were sub-aliquoted and stored at -20°C for consistency). This allowed us to transform the logarithmic cycle threshold (CT) values into linear values. Reactions were performed in duplicate or, if samples did not amplify well or if the correlation between the runs was poor, in triplicate. All PCRs were normalized by the averaged expression of three housekeeping genes ACTB, B2M and TMED10 run in triplicate. Primer sequences are listed in the Supplement.

OpenArray transcript expression. Two micrograms of total RNA were reverse transcribed using the High Capacity cDNA RT Kit (Applied Biosystems) with random hexamer primers. All cDNA was analyzed on the Cancer Pathways OpenArray system (BioTrove, Woburn, MA) using the Fast Start DNA SYBR Green kit (Roche, Nutley, NJ). Four cDNA samples were tested simultaneously per plate, with 16 samples per run. CT values were transformed into linear values by calculating 1.735 ^ (32 - CT). Values were normalized by the mean of 18 housekeeping genes. The expression values from the OpenArray platform and the qPCR (SYBR) were mean-normalized to allow for comparison across the two platforms.

Cluster analysis. Unsupervised hierarchical clustering was performed using Spearman rank correlation as similarity metric and centroid linkage as clustering method. PCR expression values were averaged between duplicate runs, mean-normalized and entered into the Cluster program (14) as log2 values. The tree was visualized using Java Treeview (15).

Unpublished Data

RESULTS AND DISCUSSION

Determination of components of variation of normal tissue. The primary goal was to identify genes that discriminate between normal breast tissue and invasive carcinoma of the breast. Since normal breast tissue was to be used as a reference, a threshold needed to be determined above which a gene would be labeled as differentially expressed. Therefore we assessed the variability in gene expression within normal breast tissue; to our knowledge, this has not preciously been well studied. Quantitative PCR (SYBR) expression analysis was performed for 18 genes on an average of 3.5 tissue slices per breast from 10 women with bilateral reduction mammaplasty. The PCRs were normalized by their median and the duplicate runs were averaged. Table 1 shows overall gene expression variability (as standard deviation) and which fraction of it is attributable to the component variabilities of woman-to-woman (averaging 64% ±9%), left-to-right breast (averaging 6% ±3%) and within-breast (averaging 30% ±9%). These percentages represent the overall magnitude of the different sources of variation as determined by ANOVA analyses. As expected, between-woman variability is greatest, twice that of within-breast variability, implying that the largest source of variation is heterogeneity among women. The smallest source of variation is the between-breast component, implying that normal material from a contralateral breast is a good surrogate for normal material from the affected breast. Approximately 30% of overall variation was explained

Unpublished Data

by variation in the molecular behavior of different biopsies of the same breast (note the assays were performed in duplicate, and the coefficient of variation (CV) of the assay was small compared to each of these components of variation and so can be ignored). Because each value was median-centered, the standard deviation also represents a surrogate for CV in the population; it is scaled to the typical expression levels. Six of the 18 genes have standard deviations above 1 (COL1A2, CTGF, GATA3, LYZ, MUC1 and WFDC2) that could be related to a spotty expression pattern (e.g. only a few cells in a tissue express the transcript). Of note, this elevated variability is unrelated to within-breast variability. For WFDC2, the most variable gene, the standard deviation is 2.73. Therefore, for subsequent cancer-to-normal ratios a threshold of the mean plus 3 standard deviations was chosen for all genes.

Database mining identifies genes with differential expression between breast cancer and normal tissues. Despite a large body of research on gene and protein expression in breast cancer, few studies include healthy controls. In those that do (16-25), the normal tissues are often not well characterized. Publications reporting expression only in breast cancer and not in healthy controls are still useful for a cancer-to-normal comparison since the data can potentially be matched with those from other sources using healthy normal tissues. These and additional expression data were compiled in a database (LevelsDB) that was then mined starting with genes contained in breast cancer data sets (4405 genes), followed by removal of genes and proteins whose subcellular localization makes the protein unlikely to be found in the blood stream by non-necrotic processes (nuclear, mitochondrial and ribosomal), leaving 3271. In a next step, housekeep-

Unpublished Data

ing genes were excluded, followed by removal of genes with high levels of expression in normal tissues of organs with strong blood contact (kidney, lungs, liver, heart and pancreas). The remaining genes and proteins were filtered for those with low expression in normal tissues. The lack of truly normal breast expression data, except for data from one normal tissue coming from a breast with cancer (26) required the use of other normal, especially epithelial tissues for subtractive comparison, contained in six datasets in LevelsDB. The last reduction step resulted in 150 genes and proteins of which 44 were likely to be secreted or membrane bound which makes them ideal as a blood marker. These were augmented by 90 genes for which literature review suggests a potential role as breast cancer markers, resulting in 134 genes (Table 2) for subsequent expression analysis by PCR. References for these additional genes are listed in the Supplement.

Expression validation results in 46 differentially expressed genes. To identify from the **134** genes those with the ability to discriminate between normal and malignant breast tissue, PCR expression analysis was performed on 93 tissues (24 invasive cancers, 38 ipsilateral normals, 3 contralateral normals, 28 tissues from breast reduction surgery) from 64 women (Supplemental Table S1). The cDNA was oligo-dT primed. A comma-delimited file with the expression data is available in the Supplement. PCR results for 8 genes were not conclusive even after 2 repeats and the genes were removed from further analysis (supplemental Table S2). Of the 126 remaining genes, 67 discriminated between the 25 cancer tissues and the 28 mammaplasty controls with ≥20% of the cancers and ≤5% of the controls above or below threshold (Table 2, and supplemental Table S2).

Unpublished Data

After completion of the PCR work, a new PCR-based technology (OpenArray by BioTrove, Woburn, MA) had come to the market that allowed for more rapid gene expression analysis (27). This technology was used to confirm the expression of a subset of the genes on a subset of the tissues. BioTrove's Cancer Pathways OpenArray plate included primer pairs specific for 606 genes associated with DNA repair, angiogenesis, cell adhesion, apoptosis, cell cycle and many genes encoding kinases. Of the 134 genes from database mining, 41 overlapped with the OpenArray panel which was suitable for confirmation of about 30% of the original results. Out of the 94 tissues that were originally tested, 13 were randomly selected from the 24 cancer tissues and likewise 9 from the 28 mammaplasty controls. Applying the same criteria as for the original set, the OpenArray analysis of the reduced set selected the same differentially expressed genes as the analysis of the original set (supplemental Table S3). Both amplifications had good correlation (supplemental Table S4) with an averaged coefficient of variation of 38% (12%-71%), even considering the differences between both methods (oligo-dT versus random priming; differences in primer sequences).

Unsupervised cluster analysis of the PCR data shows that 46 of the 67 genes have the power to separate cancer tissues from controls, thus confirming the validity of our approach (Figure 1, red and green bars in the left). These genes warrant further investigation as potential blood markers for early detection.

The mammaplasty control patients, seen in Los Angeles, were on average 15 years younger than the patients with tumor or ipsi- and contralateral normal tissue, seen in Seattle. While differences in institution and age could potentially introduce bias, the interspersing of the normal tissues in the control cluster suggests that this is not the case

Unpublished Data

(Figure 1). This can be attributed in part to the strict adherence to identical specimen collection protocols at both sites. No clustering behavior was found based on other factors listed in the supplemental Table S1, including mammography (BI-RADS) score and breast density at the last mammogram before diagnosis, lymph node positivity, tumor size, number of foci, stage and grade.

Comparison to previously published results. Comparing the present gene expression results to those previously published is difficult because prior studies rarely used healthy normal controls. While the terms “over-” and “under-expression” are common in the breast cancer literature, they most often refer to expression of one cancer state relative to another or to a cell line, and not relative to a healthy normal control. The three breast cancer publications that used mammaplasty tissue as controls confirm the expression pattern of the metalloproteinases (28), YWHAZ (11), ERBB3 and ERBB4 (29). Furthermore COL1A1 and COL1A2 over-expression was also seen in a meta-analysis of 13 publications comparing breast cancer to largely undefined and probably ipsilateral normal tissue (30). This gives credence to the observed expression pattern of the remaining genes.

Identification of additional differentially expressed genes. The OpenArray Cancer Pathways chip contained 606 genes of which 41 were used for confirmation of the PCR results. Unsupervised cluster analysis of the remaining 565 genes in the 13 cancers and 9 control tissue results in a clear distinction between over- and under-expressed genes. The OpenArray PCR was not duplicated and thus the results have greater error mar-

Unpublished Data

gins. Therefore more stringent filtering conditions were applied than for the original PCR: genes with differential expression in more than 30% of the tumors at a tumor-to-normal ratio of 1.2 were removed from the dataset. Of the resulting 102 genes, 88% were found to be under-expressed in the tumor tissues (see supplemental Figure S1), as indicated by the negative %CV numbers in Table 3. Once confirmed in their expression, it is likely that some of these 102 genes will be added to the 46 potential markers.

The high number of under-expressed genes contrasts sharply with above results where over- and under-expressed genes are equally represented. The Cancer Pathways genes had been selected based on general cancer literature which includes a large number of within-cancer and cell line experiments. Our data mining on the other hand focused on breast cancer, normal-to-cancer differences and extracellular expression. Hence, the former contains a larger number of intracellular, regulatory proteins. Interestingly, the differentially expressed genes in both sets are enriched for connective tissue genes, suggesting that alteration in the composition of the connective tissue is an important factor in cancer formation.

Histologically normal tissues from an affected breast can demonstrate molecular predisposition to cancer. Unsupervised cluster analysis of the original PCR data placed 7 of the 38 ipsilateral normal tissues in the cancer cluster (Figure 1). The difference between these and the remaining 31 ipsilateral tissues cannot be correlated with tissue or patient characteristics (supplemental Table S1). Because BRCA status was not recorded, the study does not address any possible link between mutation and a cancer-like gene expression pattern in ipsilateral normal tissue. Another explanation is a positional effect

Unpublished Data

related to distance between lesion and the site of normal tissue collection or that index lesion and ipsilateral normal tissue come from the same lobe (31). Consequently, a normal tissue from an unaffected contralateral breast should display a normal-like gene expression pattern. Indeed, of the three contralateral normal tissues, the two coming from a breast without evidence of cancer were found in the normal cluster and the third, from a breast with malignancy, grouped with the cancers.

Tripathi *et al.* compared normal tissue from mammaplasty to ipsilateral normal and breast tissue with *in situ* disease. They found that global gene expression abnormalities exist in both normal epithelium of breast cancer patients and early cancers (11). The results presented here go one step further by including same-patient invasive tissues. This leads to the conclusion that ipsilateral normal tissues with cancer-like gene expression are molecularly predisposed to cancer. To validate these findings, BRCA status of the patient and positional information of the tissue pieces harvested from a breast would need to be recorded. Our tissue collection protocol has been altered accordingly.

Identification of novel prognostic markers. Unsupervised cluster analysis of the original PCR data resulted in a cancer cluster with two sub-clusters, one enriched for patients with cancers of the luminal subtype and one of the basal subtype, as defined by hormone receptor and HER2 expression (32) (Figure 1). The composition of these two sub-clusters is summarized in supplemental Table S5. The luminal-like sub-cluster is defined by over-expression of the luminal markers ESR1, PGR and GATA3 (33-35) in all of its tissues and by the over-expression of CTGF, MMP2, AR, CFB, CD44, EPOR, CDKN1B, ETAA1, FGFR2, TNFRSF10B, ERBB4, SCUBE2, FOXA1 and MUC1 in tissues from

Unpublished Data

cancers with lobular or mixed ductal-lobular histology. Except for AR (36) and ESR1 (37), none of these genes can be linked to lobular cancer histology, in particular the comparison by Zhao *et al*. of ductal and lobular carcinomas (38). The role of GATA3 for the maintenance of the luminal phenotype has been reviewed by Tlsty, particularly the correlation of low expression of GATA3 and low estrogen receptor alpha (39) which the present data confirm. The basal-like sub-cluster, characterized by the under-expression of these genes, is enriched for triple-negative (hormone receptor and HER2-negative) cancers and contains all cancer tissues of patients that are deceased (black dots) or have recurred (orange dots). Lobular breast carcinomas are known to be associated with better survival than ductal carcinomas (37, 40) and triple-negative breast cancers have been associated with poor prognosis (41). Also, in a meta-analysis of published breast cancer cDNA data, low GATA3 expression is linked with poor clinical outcome (42). The difference between these cancer sub-clusters could therefore be attributed to the aggressiveness of the disease. While some of these genes have been linked to the basal subtype (16, 19) and some are now being used to predict disease outcome, including SCUBE2 in Oncotype DX (43) and Mammaprint (44), the majority of them may constitute a novel group of genes that predict outcome and/or inform treatment.

Interestingly, the under-expressed genes that define the basal-like cluster are under-expressed in most controls, suggesting that these aggressive cancers may be difficult to distinguish from normal breast tissue at the molecular level.

Conclusions. Our results suggest that many of the genes commonly attributed to cancer pathways are expressed at lower levels in breast cancers than in normal breast tissue,

Unpublished Data

confirming and further extending results by Tripathi *et al.* (11). Furthermore, the genes that predict aggressive phenotype in between-cancer comparisons are not differentially expressed between aggressive cancers and healthy controls. If serum assays commonly measure the increase of a marker rather than its absence in cancer, our findings would help explain the current lack of suitable blood markers for breast cancer, particularly in patients with poor prognosis malignancies.

In spite of these shortcomings, our work resulted in the identification of a number of differentially expressed genes, including 12 related to aggressive disease, a minimum of 46 discriminating between cancer and controls, of which some (MMP12, S100A7 and SPP1) are over-expressed in aggressive cancer. Those coding for proteins that are readily shed may be of greatest interest for serum marker evaluation, and markers that are over-expressed but not shed may be the most attractive for tumor-specific localization, including prognosis.

ACKNOWLEDGEMENTS

Work was supported by DOD/CDMRP DAMD17-02-1-0691 and by NCI/Avon P50 CA083636. We would like to thank the Swedish Cancer Institute surgeons Patricia Dawson, Claire Buchanan, James Hanson, Christine Lee and Darlene Barr and the pathologists at Cellnetix, in particular Ron Tickman and Nan Ping Wang, Christopher Miller

Unpublished Data

and Anthony Blau for providing the EPO and EPOR primers and Max Drescher for his help with the PCR.

Unpublished Data

Bibliography

1. Fletcher RH. Should all people over the age of 50 have regular fecal occult-blood tests? If it works, why not do it? New England Journal of Medicine 1998; 338: 1153-4; discussion 4-5.
2. Tabár L, Vitak B, Chen HH, Prevost TC, Duffy SW. Update of the Swedish Two-County Trial of breast cancer screening: histologic grade-specific and age-specific results. Swiss Surgery 1999; 5: 199-204.
3. Gøtzsche PC, Nielsen M. Screening for breast cancer with mammography. Cochrane Database Syst Rev 2006; 18: CD001877.
4. Fletcher SW, Black W, Harris R, Rimer BK, Shapiro S. Report of the International Workshop on Screening for Breast Cancer [see comments]. Journal of the National Cancer Institute 1993; 85: 1644-56.
5. Christiansen CL, Wang F, Barton MB, et al. Predicting the cumulative risk of false-positive mammograms. Journal of the National Cancer Institute 2000; 92: 1657-66.
6. Duffy MJ. Serum tumor markers in breast cancer: are they of clinical value? Clin Chem 2006; 52: 345-51.
7. Schummer M, Ng W, Bumgarner R, et al. Comparative hybridization of an array of 21,500 ovarian cDNAs for the discovery of genes overexpressed in ovarian carcinomas. Gene 1999; 238: 375-85.
8. Pepe MS, Etzioni R, Feng Z, et al. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst 2001; 93: 1054-61.

Unpublished Data

9. Hellström I, Raycraft J, Hayden-Ledbetter M, et al. The HE4 (WFDC2) protein is a biomarker for ovarian carcinoma. Cancer Res 2003; 63: 3695-700.

10. Scholler N, Crawford M, Sato A, et al. Bead-Based ELISA for Validation of Ovarian Cancer Early Detection Markers. Clin Cancer Res 2006; 12: 2117-24.

11. Tripathi A, King C, de la Morenas A, et al. Gene expression abnormalities in histologically normal breast epithelium of breast cancer patients. Int J Cancer 2008; 122: 1557-66.

12. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res 2005; 33: D54-8.

13. Wain HM, Bruford EA, Lovering RC, Lush MJ, Wright MW, Povey S. Guidelines for human gene nomenclature. Genomics 2002; 79: 464-70.

14. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proceedings of the National Academy of Sciences of the United States of America 1998; 95: 14863-8.

15. Saldanha AJ. Java Treeview--extensible visualization of microarray data. Bioinformatics 2004; 20: 3246-8.

16. Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000; 406: 747-52.

17. Dressman MA, Walz TM, Lavedan C, et al. Genes that co-cluster with estrogen receptor alpha in microarray analysis of breast biopsies. Pharmacogenomics J 2001; 1: 135-41.

Unpublished Data

18. Sørlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001; 98: 10869-74.

19. Sørlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003; 100: 8418-23.

20. Zhao H, Langerod A, Ji Y, et al. Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. Mol Biol Cell 2004; 15: 2523-36.

21. Amatschek S, Koenig U, Auer H, et al. Tissue-wide expression profiling using cDNA subtraction and microarrays to identify tumor-specific genes. Cancer Res 2004; 64: 844-56.

22. Abba MC, Drake JA, Hawkins KA, et al. Transcriptomic changes in human breast cancer progression as determined by serial analysis of gene expression. Breast Cancer Res 2004; 6: R499-513.

23. Turashvili G, Bouchal J, Baumforth K, et al. Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis. BMC Cancer 2007; 7: 55.

24. Grigoriadis A, Mackay A, Reis-Filho JS, et al. Establishment of the epithelial-specific transcriptome of normal and malignant human breast cells based on MPSS and array expression data. Breast Cancer Res 2006; 8: R56.

25. Covell DG, Wallqvist A, Rabow AA, Thanki N. Molecular classification of cancer: unsupervised self-organizing map analysis of gene expression microarray data. Mol Cancer Ther 2003; 2: 317-32.

Unpublished Data

26. Shyamsundar R, Kim YH, Higgins JP, et al. A DNA microarray survey of gene expression in normal human tissues. Genome Biol 2005; 6: R22.

27. Morrison T, Hurley J, Garcia J, et al. Nanoliter high throughput quantitative PCR. Nucleic Acids Res 2006; 34: e123.

28. Köhrmann A, Kammerer U, Kapp M, Dietl J, Anacker J. Expression of matrix metalloproteinases (MMPs) in primary human breast cancer and breast cancer cell lines: New findings and review of the literature. BMC Cancer 2009; 9: 188.

29. Bièche I, Lerebours F, Tozlu S, Espie M, Marty M, Lidereau R. Molecular profiling of inflammatory breast cancer: identification of a poor-prognosis gene expression signature. Clin Cancer Res 2004; 10: 6789-95.

30. Srour N, Reymond MA, Steinert R. Lost in translation? A systematic database of gene expression in breast cancer. Pathobiology 2008; 75: 112-8.

31. Tot T. The theory of the sick breast lobe and the possible consequences. Int J Surg Pathol 2007; 15: 369-75.

32. Charpin C, Giusiano S, Secq V, et al. Quantitative immunocytochemical profile to predict early outcome of disease in triple-negative breast carcinomas. Int J Oncol 2009; 34: 983-93.

33. Jacquemier J, Padovani L, Rabayrol L, et al. Typical medullary breast carcinomas have a basal/myoepithelial phenotype. J Pathol 2005; 207: 260-8.

34. Nielsen TO, Hsu FD, Jensen K, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res 2004; 10: 5367-74.

Unpublished Data

35. Fabre-Lafay S, Monville F, Garrido-Urbani S, et al. Nectin-4 is a new histological and serological tumor associated marker for breast cancer. BMC Cancer 2007; 7: 73.

36. Narita D, Raica M, Suciu C, Cimpean A, Anghel A. Immunohistochemical expression of androgen receptor and prostate-specific antigen in breast cancer. Folia Histochem Cytobiol 2006; 44: 165-72.

37. Pestalozzi BC, Zahrieh D, Mallon E, et al. Distinct clinical and prognostic features of infiltrating lobular carcinoma of the breast: combined results of 15 International Breast Cancer Study Group clinical trials. J Clin Oncol 2008; 26: 3006-14.

38. Zhao H, Langerød A, Ji Y, et al. Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. Mol Biol Cell 2004; 15: 2523-36.

39. Tlsty TD. Luminal cells GATA have it. Nat Cell Biol 2007; 9: 135-6.

40. Dian D, Herold H, Mylonas I, et al. Survival analysis between patients with invasive ductal and invasive lobular breast cancer. Arch Gynecol Obstet 2009; 279: 23-8.

41. Sivridis E, Stamos C, Fiska A, Nikolettos N, Koukourakis MI, Giatromanolaki A. c-erbB-2 and the "triple-state" in early breast carcinomas. Med Oncol 2009; 23: 23.

42. Oh DS, Troester MA, Usary J, et al. Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. J Clin Oncol 2006; 24: 1656-64.

43. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004; 351: 2817-26.

Unpublished Data

44. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002; 415: 530-6.

Unpublished Data

TABLES

Table 1. Variability in gene expression

Marker	StDev	Between women	Between breasts within woman	Within breast
ASPN	0.88	75%	3%	21%
CAV1	0.89	60%	2%	38%
CFB	0.74	46%	8%	46%
COL1A2	1.13	76%	10%	15%
CTGF	1.07	67%	6%	28%
GATA3	1.61	60%	4%	36%
LETMD1	0.46	79%	9%	11%
MGST1	0.94	59%	4%	37%
LYZ	1.08	66%	8%	23%
MMP2	0.63	50%	10%	39%
MUC1	2.34	62%	4%	34%
SPARC	0.71	74%	3%	22%
SUMF2	0.39	55%	7%	35%
TIMP1	0.69	65%	9%	26%
TIMP2	0.49	63%	7%	29%
TIMP3	0.60	67%	3%	30%
WFDC2	2.73	61%	4%	35%
YWHAZ	0.37	64%	5%	30%
	Mean	**64%**	**6%**	**30%**
	StDev	9%	3%	9%

Variability in gene expression across all tissue pieces (StDev) and as percentage by each of the three fractions. The last two rows display the mean and standard deviation of the 18 genes.

Table 2. 71 of 126 genes discriminate cancers from controls

Gene Symbol	Result	Gene Symbol	Result	Gene Symbol	Result
ADAM12*		**EPOR**	∇	**PEBP4**	
AGR2*		**ERBB3***		**PGR**	∇
AKT1		**ERBB4**	∇	**PIK3CA**	
AMBP*		**ESR1**	Δ	**PIP**	
ANGPT2*	∇	**ETAA1***	∇	**PLAUR**	Δ
APOL1*	∇	**FGFR2***	∇	**PRLR***	
AR	∇	**FN1***	Δ	**PROCR**	∇
ASPN*	Δ	**FOXA1**	Δ	**PSMA5***	

Unpublished Data

Table 2. 71 of 126 genes discriminate cancers from controls

Gene Symbol	Result	Gene Symbol	Result	Gene Symbol	Result
BGN*	Δ	GATA3	Δ	PTPN1*	
BIRC5	Δ	GDF15		PVRL4	
BRCA2		HOXB7	∇	S100A7	Δ
BRMS1		IFIT1*		S100B*	∇
BUB1	Δ	IGF2		SCGB2A1	
C18orf8*		KRT20		SCUBE2*	∇
CALB2		KRT7		SDC1	
CAV1	∇	LCN2		SFRP1	∇
CCNE1	Δ	LETMD1	∇	SFRP2*	
CD274	Δ	LPAR3		SNIP	
CD44	∇	LRRC15*	Δ	SPARC	∇
CDH1		LTF*		SPP1	Δ
CDKN1B*	∇	LYZ*		STC2*	
CDX2		MGST1	∇	SUMF2*	
CFB*	Δ	MIF		THBS2	
COL11A1*	Δ	MMP1	Δ	TIMP1	Δ
COL1A1*	Δ	MMP10	Δ	TIMP2	∇
COL1A2*	Δ	MMP11	Δ	TIMP3	∇
COL3A1*		MMP12	Δ	TIMP4	∇
COL5A1*	Δ	MMP13	Δ	TK1	Δ
COL5A2*	∇	MMP14		TM9SF2*	
COL6A3*		MMP16		TNFRSF10B	∇
COL8A1*		MMP2	∇	TNN	∇
COMP*	Δ	MMP20		TOP2A	Δ
CSNK2A1		MMP3		TP53	
CTGF	∇	MMP7	∇	TRPS1	
CTHRC1*	Δ	MMP8		TTF1	
CYP4B1*	∇	MMP9		VCAN*	Δ
CYR61	∇	MSLN		VEGFA	
DEFA1		MUC1	Δ	VTCN1	
DEFA3		NES	∇	WFDC2	Δ
ECM1*		OAS1*		WT1	Δ
EGFR	∇	OAS2*	Δ	XBP1	
EPO		PEBP1	∇	YWHAZ	

126 genes found by mining of expression data and/or LevelsDB (asterisk). Thresholds for cancers and controls were determined by expression in the 28 normal mammaplasty tissues as mean +3 SD for genes with over-expression in the cancers and as below the minimum for genes with under-expression in the cancers. Over- (Δ) or under- (∇) expression in cancer tissue with ≥20% of the cancers and ≤5% of the controls above or below threshold.

Unpublished Data

Table 3. Genes resulting from the OpenArray analysis

CSF1	-100%	HADHA	-77%	TP53I3	-62%	SLC2A3	-38%
EGR1	-100%	MCC	-77%	ANPEP	-54%	SMPD1	-38%
FLT1	-100%	RELA	-77%	BAG1	-54%	TGFBI	-38%
FOS	-100%	BTG2	-69%	ILK	-54%	BNIP3	-31%
NID1	-100%	CNBP	-69%	ING1	-54%	CBLB	-31%
SEPP1	-100%	DHX8	-69%	PECAM1	-54%	DEGS1	-31%
SRPX	-100%	EPHA2	-69%	PIR	-54%	EGLN1	-31%
TGFBR2	-100%	GNB2L1	-69%	RIPK1	-54%	ETV6	-31%
TGFBR3	-100%	IGFBP4	-69%	SFRS7	-54%	FOSL2	-31%
TIE1	-100%	NDRG1	-69%	TSG101	-54%	LDHA	-31%
VIM	-100%	PAQR3	-69%	CAPNS1	-46%	NR1D1	-31%
HYAL1	-92%	PEA15	-69%	CHPT1	-46%	PRKCD	-31%
PPARG	-92%	PFDN5	-69%	EIF5	-46%	PRNP	-31%
RAB5A	-92%	RAF1	-69%	GTF2I	-46%	SORT1	-31%
SEMA3C	-92%	RAP1A	-69%	JAK1	-46%	TRADD	-31%
SPRY2	-92%	SKI	-69%	MDM2	-46%	EVL	31%
CCND3	-85%	SP1	-69%	MLLT10	-46%	HSPB1	31%
CDC42BPA	-85%	STK3	-69%	SELENBP1	-46%	KIF3B	38%
CIRBP	-85%	CSF1R	-62%	ATP5B	-38%	PKM2	38%
FOXO1	-85%	GAS6	-62%	AXL	-38%	RFC4	38%
ITGB3	-85%	NF2	-62%	CTNNA1	-38%	RARA	46%
PTEN	-85%	PECI	-62%	DCN	-38%	RAD21	54%
RHOB	-85%	PRKCE	-62%	EXT1	-38%	PRC1	77%
TYRO3	-85%	STAT3	-62%	HRB	-38%	SKIL	77%
ABL1	-77%	TAF1	-62%	PPP2R5A	-38%		
CD59	-77%	TJP1	-62%	PRKD2	-38%		

List of the 102 genes from the OpenArray analysis and percentage of tumor tissues they were differentially expressed in. Negative numbers indicate under-expression.

FIGURE LEGEND

Figure 1. Unsupervised hierarchical clustering of 93 tissues (24 invasive cancers, 38 ipsilateral normal, 3 contralateral normal, 28 normal tissues from reduction mammaplasty) from 64 patients and 67 genes that discriminate between invasive tissues and mammaplasty normal tissues (red and green dots: over- and under-expression by PCR). Columns: tissues form two distinct clusters (indicated below the figure). Rows:

Unpublished Data

genes form a cancer and a normal cluster, the latter being divided in one with underexpression in all cancer tissues (left, green line) and one with mixed expression (orange-blue line). Luminal-like and basal-like clusters are indicated above the figure. The part of the heat map driving the luminal-like cluster is boxed (blue: luminal-like genes, turquoise: lobular tissues). Tissues from deceased or recurred patients have a black or orange dot above the tissue descriptor which has the following abbreviated components: PatientNo – Diagnosis (IDC=invasive ductal carcinoma, ILC=invasive lobular carcinoma, MET=metaplastic carcinoma, MUC=mucinous carcinoma, NML=normal) – Stage – Grade TissueNo – Description (CA=cancer, NM=normal mammaplasty, NI=normal ipsilateral, NC=normal contralateral) BI-RADS Density Subtype (LUM=luminal, BAS=basal HER2). The tissue descriptors are shaded as follows: orange=lobular cancers, pink=other cancers, green=ipsilateral normals, blue=contralateral normals, purple=mammaplasty normals. Heat map: Red=up-regulation, green=down-regulation, grey=missing or zero value. The lines below the heat map connect tissues from the same patient.

Unpublished Data

STEM CELLS®

TRANSLATIONAL AND CLINICAL RESEARCH

Evaluating Erythropoietin-Associated Tumor Progression Using Archival Tissues from a Phase III Clinical Trial

CHRIS P. MILLER,[a] KIMBERLY A. LOWE,[b] KARINE VALLIANT-SAUNDERS,[a] JORINGEL F. KAISER,[a] DOMINIK MATTERN,[c] NICOLE URBAN,[b] MICHAEL HENKE,[c] C. ANTHONY BLAU[a]

[a]Department of Medicine/Hematology and the Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, Washington, USA; [b]Fred Hutchinson Cancer Research Center, Seattle, Washington, USA; [c]Department of Radiooncology, University Clinic, Freiburg, Germany

Key Words. Erythropoietin • Erythropoietin receptor • Erythropoiesis stimulating agents • Growth factors • Tumor progression

ABSTRACT

Despite the prevalence of anemia in cancer, recombinant erythropoietin (Epo) has declined in use because of recent Phase III trials showing more rapid cancer progression and reduced survival in subjects randomized to Epo. Since Epo receptor (EpoR), Jak2, and Hsp70 are well-characterized mediators of Epo signaling in erythroid cells, we hypothesized that Epo might be especially harmful in patients whose tumors express high levels of these effectors. Because of the insensitivity of immunohistochemistry for detecting low level *EpoR* protein, we developed assays to measure levels of *EpoR*, *Jak2* and *Hsp70* mRNA in formalin-fixed paraffin-embedded (FFPE) tumors. We tested 23 archival breast tumors as well as 136 archival head and neck cancers from ENHANCE, a Phase III trial of 351 patients randomized to Epo versus placebo concomitant with radiotherapy following complete resection, partial resection, or no resection of tumor. *EpoR, Jak2, and Hsp70* mRNA levels varied >30-fold, >12-fold, and >13-fold across the breast cancers, and >30-fold, >40-fold, and >30-fold across the head and neck cancers, respectively. Locoregional progression-free survival (LPFS) did not differ among patients whose head and neck cancers expressed above- versus below-median levels of *EpoR*, *Jak2* or *Hsp70*, except in the subgroup of patients with unresected tumors ($n = 28$), where above-median *EpoR*, above-median *Jak2*, and below-median *Hsp70* mRNA levels were all associated with significantly poorer LPFS. Our results provide a framework for exploring the relationship between Epo, cancer progression, and survival using archival tumors from other Phase III clinical trials. STEM CELLS *2009;27:2353–2361*

Disclosure of potential conflicts of interest is found at the end of this article.

INTRODUCTION

Anemia is common in cancer patients and likely represents an independent poor prognostic factor for survival [1]. Safety concerns associated with transfused blood elevated erythropoietin (Epo) to a mainstay treatment in oncology. However, recent Phase III clinical trials testing new uses for Epo, including targeting higher hemoglobin levels and treating anemia not caused by chemotherapy, showed that Epo reduced cancer survival times. Venous thromboembolism is a well documented risk of Epo [2], however the adverse outcomes in these trials were attributed mainly to accelerated tumor progression [3–7].

Whether Epo can indeed stimulate cancer progression is the subject of an intense controversy [8], and preclinical models have generated conflicting results (reviewed by Arcasoy [9]). Central to the controversy is whether tumor progression reflects an "off-target" interaction between Epo and Epo-responsive tumor cells and/or tumor blood vessels. At issue is whether tumors (or tumor blood vessels) can expropriate signaling pathways known to confer Epo responsiveness in erythroid cells. Epo receptor (*EpoR*) mRNA and protein are detectable in tumor cells, albeit at levels much lower than in erythroid cells [10, 11]. Notably, a recent study showed that a neuroblastoma cell line expressing fewer than 50 Epo binding sites per cell can still be protected from apoptosis in response to Epo [12]. Thus, the pertinent unanswered question is whether even low-level expression of EpoR or other effectors of Epo-signaling can promote cancer progression in patients treated with Epo.

A direct approach to examining this issue would be to characterize archival tumor specimens from patients who had enrolled in Phase III clinical trials of Epo versus placebo, testing whether randomization to Epo was especially harmful in those patients whose tumors expressed higher levels of EpoR and/or downstream effectors of Epo signaling. A previous study employing this approach characterized 154 archival tumors

Author contributions: C.P.M.: conception and design, collection and assembly of data, data analysis and interpretation, manuscript writing, final approval of manuscript; K.A.L. and N.U.: data analysis and interpretation, final approval of manuscript; K.V.-S. and J.F.K.: collection and assembly of data; D.M.: provision of study materials; M.H.: conception and design, provision of study materials, data analysis and interpretation, manuscript writing, final approval of manuscript; C.A.B.: conception and design, financial support, assembly of data, data analysis and interpretation, manuscript writing, final approval of manuscript.

Correspondence: C. Anthony Blau, M.D., University of Washington, Box 358056, 815 Mercer Street, Seattle, WA 98109. Telephone: 206-685-6873; Fax: 206-221-0531; e-mail: tblau@u.washington.edu Received January 8, 2009; accepted for publication June 8, 2009; first published online in STEM CELLS *EXPRESS* June 18, 2009; available online without subscription through the open access option. © AlphaMed Press 1066-5099/2009/$30.00/0 doi: 10.1002/stem.156

from ENHANCE, a Phase III trial of 351 patients randomized to Epo versus placebo concomitant with radiotherapy following complete resection, partial resection, or no resection of head and neck cancer [13]. Tumors were evaluated using a commercially available polyclonal antibody raised against a human EpoR peptide (C20), that also cross-reacts with non-EpoR proteins, including heat shock protein 70 (Hsp70) family members [14]. A significant association between Epo assignment and reduced LPFS was observed among patients with C20-positive tumors ($p = .003$, $n = 104$) that was not observed in patients with C20-negative tumors. However, the aforementioned cross-reactivity between C20 and non-EpoR proteins obscured the interpretation of this finding.

Because of the inadequacy of reagents for detecting low-level EpoR protein in archival tumors, we measured mRNA. Most clinical tumor specimens are formalin-fixed and paraffin-embedded (FFPE), causing RNA degradation. We therefore developed methods to measure mRNA levels of *EpoR* and 16 other genes from FFPE tumors. To test whether the adverse effects of Epo might be mediated by increased expression of other genes implicated in Epo-responsiveness, we included *Csf2rb, Jak2, and Hsp70*. *Csf2rb* encodes the common beta receptor (βcR), a shared signaling subunit for several cytokine receptors, that has been suggested to enhance Epo signaling in nonerythroid cells [15]. Jak2 is a tyrosine kinase that is an essential mediator of Epo signaling in erythroid cells [16], facilitates cell surface EpoR expression [17], and is also implicated in Epo-mediated neuroprotection [18]. Hsp70 family members are encoded by eight *Hspa* genes, perform essential roles in protein folding, transport, and degradation [19], and promote cancer cell survival [20]. *Hspa1a* and *Hspa1b* encode proteins with one amino acid difference, collectively referred to as the major stress inducible Hsp70. In differentiating erythroid cells, Hsp70 accumulates in the nucleus in response to Epo, where it shields the transcription factor Gata-1 from caspase-3-mediated degradation [21]. Additional markers were included to test whether the adverse effects of Epo might depend on vascular endothelial cell representation (*Cdh5, Pecam1, Vegfa*), tumor squamous epithelial cell representation (*Krt5*) [22], or cancer stem cells (*Cd44*) [23], since a recent study suggested that Epo may increase the self-renewal capacity of $CD44^+$ breast cancer-initiating cells [24]. We also measured transcripts for *Epo* itself, and seven control genes for normalization (see below). Our results provide a framework for investigating Epo-induced tumor progression.

METHODS

Cell Lines

All cancer cell lines have been previously described. To prepare Ba/F3-h*EpoR* cells, Ba/F3 cells [25] were electroporated with pcDNA3.1-h*EpoR* encoding a human *EpoR* cDNA (a gift from Joseph Prchal, University of Utah), selected in 1 mg/ml Geneticin (Invitrogen, Carlsbad, CA, http://www.invitrogen.com), and maintained in 1U/ml epoetin alfa (Procrit, Ortho Biotech, Bridgewater, NJ, http://www.orthobiotech.com). COS-h*EpoR* cells were prepared by transfecting COS cells with pcDNA3.1-h*EpoR* using Lipofectamine 2000 (Invitrogen) and were collected 48 hours after transfection. AT-2 cells were provided by Janet Rowley (University of Chicago) and ASE2 cells were provided by Chugai Pharmaceuticals (Japan).

Immunohistochemistry

FFPE cell pellets were sectioned (6 micron) and slides were deparaffinized and rehydrated through a graded ethanol series. Endogenous peroxidase activity was blocked using 0.3% hydrogen peroxide for 8 minutes, and endogenous biotin sites were blocked using the Avidin/Biotin Blocking Kit (Dako, Glostrup, Denmark, http://www.dako.com). Sections were then incubated with a polyclonal goat anti-EpoR antibody (ab10653, Abcam, Cambridge, MA, http://www.abcam.com) for 60 minutes. Primary antibodies were detected using a biotinylated antigoat secondary antibody (Jackson ImmunoResearch, West Grove, PA, http://www.jacksonimmuno.com) for 30 minutes followed by visualization using the Vector Elite ABC system (Vector Laboratories, Burlingame, CA, http://www.vectorlabs.com). Staining was visualized with 3,3′-diaminobenzidine for 7 minutes, and the sections were counterstained with hematoxylin for 2 minutes. Concentration matched isotype controls (Jackson ImmunoResearch) were run for each cell sample.

Flow Cytometric Detection of Cell Surface EpoR

Adherent cell lines were lifted for 15 minutes using 0.02% ethylenediaminetetraacetic acid in phosphate buffered saline (PBS), washed with PBS, and filtered through a 70 μm strainer. Cells were blocked for 15 minutes at room temperature in fluorescence-activated cell sorting (FACS) buffer (PBS, 0.1% bovine serum albumin (BSA), 0.02% sodium azide) containing 250 μg/ml human immunoglobulin G (IgG) (Sigma-Aldrich, St. Louis, MO, http://www.sigmaaldrich.com). A murine monoclonal antihuman EpoR-phycoerythrin (PE) antibody (FAB307P, R&D Systems, Minneapolis, MN, http://www.rndsystems.com) was then added to 5 μg/ml and cells were incubated for 30 minutes on ice. Cells were also stained with two different murine $IgG2_b$-PE isotype control antibodies (IC0041P, R&D Systems or 555058, BD Biosciences, Franklin Lakes, NJ, http://www.bdbiosciences.com). After staining, cells were washed, resuspended in FACS buffer, and analyzed by flow cytometry (FACS-Canto, BD Biosciences). Dead cells were excluded from analyses of adherent cells by inclusion of 3.75 μg/ml 7-aminoactinomycin D. For determination of EpoR staining relative to each isotype control, the mean fluorescence value obtained for three replicate isotype control staining reactions was subtracted from the mean fluorescence value obtained for three replicate anti-EpoR staining reactions.

Signal Transducer and Activator of Transcription 5 Phosphorylation

REH (acute lymphoblastic leukemia) and U266 (myeloma) cells were washed twice, starved for 5 hours in Roswell Park Memorial Institute media (RPMI), 0.5% BSA, and stimulated for 15 minutes with either 10 U/ml epoetin alfa (Procrit, Ortho Biotech), or vehicle (Procrit buffer: 2.5 mg/ml human albumin, 1.3 mg/ml sodium citrate, 8.2 mg/ml sodium chloride, 0.11 mg/ml citric acid, 1% benzyl alcohol). For Epo antagonist control reactions, a 225 amino acid recombinant soluble human EpoR extracellular domain (R&D Systems) was added to 2.25 μg/ml. Cells were then fixed in 2% paraformaldehyde (10 minutes, 37°C), washed, permeabilized with 90% methanol in PBS (30 minutes, 4°C), washed, resuspended in FACS buffer, and stained for 20 minutes at room temperature with an Alexa Fluor 647 anti-phospho- signal transducer and activator of transcription 5 (STAT5) phosphotyrosine 464 (PY464) antibody (1:5 dilution, 612599, BD Biosciences). Cells were then washed, resuspended in FACS buffer, and analyzed by flow cytometry (FACS-Canto, BD Biosciences).

Tumor Samples

Permission was obtained from the University of Washington Institutional Review Board to study primary tumors. Breast tumors were from an established repository (Department of Defense grant DAMD 17-02-1-0691) that stores tissues donated by women undergoing surgery for breast cancer (invasive cancer or in situ disease) as FFPE tissue and as snap-frozen tissue. FFPE head and neck tumors were obtained from the local Pathology Department and from ENHANCE, a previously reported [3] multicenter Phase III trial of epoetin beta in 351 patients receiving radiotherapy for head and neck cancer. All ENHANCE samples

were among 154 tumors previously examined using the polyclonal C20 anti-EpoR antibody (Santa Cruz Biotechnology, Santa Cruz, CA, http://www.scbt.com) [13]. One hundred thirty-six of the 154 tumors were sent as four micron FFPE sections on deidentified, coded glass slides to the University of Washington for mRNA analysis by an investigator blinded to clinical outcomes.

ENHANCE Trial Design

Patient selection, treatment, follow-up, evaluation, and baseline characteristics were described previously [3, 13]. Briefly, the main inclusion criteria were squamous cell carcinomas of the head and neck, scheduled definitive or postoperative radiotherapy, and a decreased blood hemoglobin (<13g/dl, men; <12g/dl, women) at randomization. Patients were randomly assigned to 300 IU/kg epoetin beta or placebo three times per week starting 10 to 14 days before radiotherapy, continuing throughout. Prior to randomization, patients were stratified by resection status: 1) complete resection; 2) incomplete resection; or 3) unresected disease. Iron (III) saccharate (200 mg) was administered intravenously once weekly to patients with <25% transferrin saturation. Epoetin beta was stopped if hemoglobin increased more than 2 g/dl within 1 week or when targets were reached (≥15 g/dl, men; ≥14 g/dl, women) and resumed when hemoglobin fell below target. Locoregional cancer control and survival was assessed at 3-month intervals by an independent oncologist blinded to treatment assignment. The primary endpoint was locoregional progression-free survival (LPFS). Locoregional progression was noted if the tumor recurred or increased by 25%. Baseline serum Epo levels were determined prior to treatment.

Quantitative Reverse Transcriptase Polymerase Chain Reaction

RNA was extracted from cancer cell lines using the RNeasy Mini Kit (Qiagen, Valencia, CA, http://www.qiagen.com) and from FFPE tumor sections using the Absolutely RNA FFPE kit (Stratagene, La Jolla, CA, http://www.stratagene.com). On-column DnaseI digestion was performed to remove genomic DNA. First strand cDNA was synthesized with random hexamer primers and Superscript III reverse transcriptase (RT) (Invitrogen), the latter omitted for no-RT control reactions. Next, cDNA targets were amplified using Taqman probes and a 7900HT thermal cycler (Applied Biosystems, ABI, Foster City, CA, http://www.appliedbiosystems.com). With the exception of certain intronless members of the *Hsp70* family and the candidate reference gene *18s*, all probes recognized exon junctions to prevent genomic DNA amplification (supporting information Table 1). Cycle threshold (Ct) values were determined with the Sequence Detection Software (ABI). A coefficient of variance <4% for triplicate Ct determinations was considered acceptable. Where indicated in the text, preamplification of cDNA was performed with the Taqman preamplification multiplex system (ABI). Preamplification uniformity (lack of bias) for each Taqman probe was tested by calculating ΔCt values for data obtained with both unamplified and preamplified cDNA, where ΔCt = mean Ct for target gene − mean Ct for reference gene. This was performed using several ENHANCE samples which contained sufficient RNA, erythroid ASE2 cells [26], and a universal human total RNA standard (Stratagene). Uniformity comparisons (ΔCt preamplified − ΔCt unamplified) were considered acceptable to a tolerance of variation of 1.5 cycles per manufacturer's instructions (ABI). Relative quantification was determined using the comparative Ct method, $2^{-\Delta CT}$ where ΔCt = mean Ct for target gene − mean Ct for reference gene. Reference gene stability was evaluated using the Genorm algorithm [27].

Statistical Analysis

The number of patients included in the analysis of LPFS for each marker depended upon the number of samples yielding sufficient RNA for quantitative reverse transcriptase polymerase chain reaction (RT-PCR). This varied for each marker as a function of the expression level of that gene. Our analyses included all available data for each marker. For LPFS analyses, patients were stratified into above-median or below/equal-median mRNA expression levels. This stratification was done separately for the total population and within each resection stratum for every gene. LPFS was evaluated with the Kaplan-Meier survival estimation. The log-rank test was implemented to test the null hypothesis that the distribution of survival times between patients treated with Epo versus placebo was equal. The STATA statistical software package was used for all analyses (version 10.0, Stata Corporation, College Station, TX, http://www.stata.com). Statistical tests were two-sided and considered significant at $p \leq 0.05$. For patients in the placebo group stratified by endogenous serum Epo, ≤11 U/l was defined as low, whereas >11 U/l was defined as high, based on a previous study [28]. Spearman's correlation coefficients were calculated for C20 staining status (measured as a dichotomous variable) versus *EpoR* or *Hsp70* mRNA levels (measured as a continuous variable). In cell line studies, Spearman's correlation coefficients were calculated for *EpoR* mRNA versus surface protein levels.

RESULTS

EpoR mRNA and Surface Protein Levels in Cancer Cell Lines

To characterize the relation between *EpoR* mRNA and cell surface protein, we tested 32 human cell lines including three high EpoR-expressing positive control cell lines: UT7EPO, ASE2, and OCIM1. For normalization of mRNA levels, we used the three most stable reference genes (*Hmbs, Hprt1, Rplp0*) among a panel of seven candidates evaluated across all cell lines as determined by the Genorm algorithm (supporting information Table 2). *EpoR* mRNA levels among the noncontrol lines ranged from 0.5 to 7.5% (mean 2.0%) of the level in UT7EPO cells (Fig. 1A). For flow cytometry, we compared the fluorescent intensity of cells stained with a monoclonal antibody directed against EpoR with the same cells stained using two different isotype control antibodies. Surface EpoR levels among the noncontrol lines ranged from 1.2 to 25.2% (mean 8.3%) of the level in UT7EPO cells (Fig. 1B). Among all cell lines tested there was a significant correlation between mRNA and surface protein ($r = .33$, $p = .03$, $n = 32$). When positive-control UT7EPO, ASE2, and OCIM1 cells were excluded, the significance of this correlation was maintained among nonadherent cells ($r = .58$, $p = .03$, $n = 11$, Fig. 1C) but no correlation was observed when our analysis was restricted to the adherent cell lines ($r = -.19$, $p = .23$, $n = 18$, Fig. 1D). Of note, our analysis of adherent cell lines ($n = 18$) required additional processing steps to generate single cell suspensions (see Methods) that were associated with significant cell death and debris (not shown). Moreover, 16 of 18 adherent cell lines produced discordant staining patterns for the two different isotype control antibodies (that is, average deviation in fluorescence values obtained for anti-EpoR antibody staining relative to fluorescence values obtained for two isotype control antibodies exceeded 5% of the mean), whereas only 5 of 11 nonadherent cell lines showed this discrepancy ($p < .01$) (Fig. 1B). The apparent lack of correlation between *EpoR* mRNA and surface protein among the adherent cell lines likely results from these technical limitations that preclude accurately estimating levels of EpoR on the cell surface, but may also be influenced by post-transcriptional regulation of EpoR in these lines. Analysis of two of the EpoR-expressing nonerythroid lines (U266 and REH) demonstrated Epo-dependent STAT5 phosphorylation (supporting information Fig. 1), consistent with previous

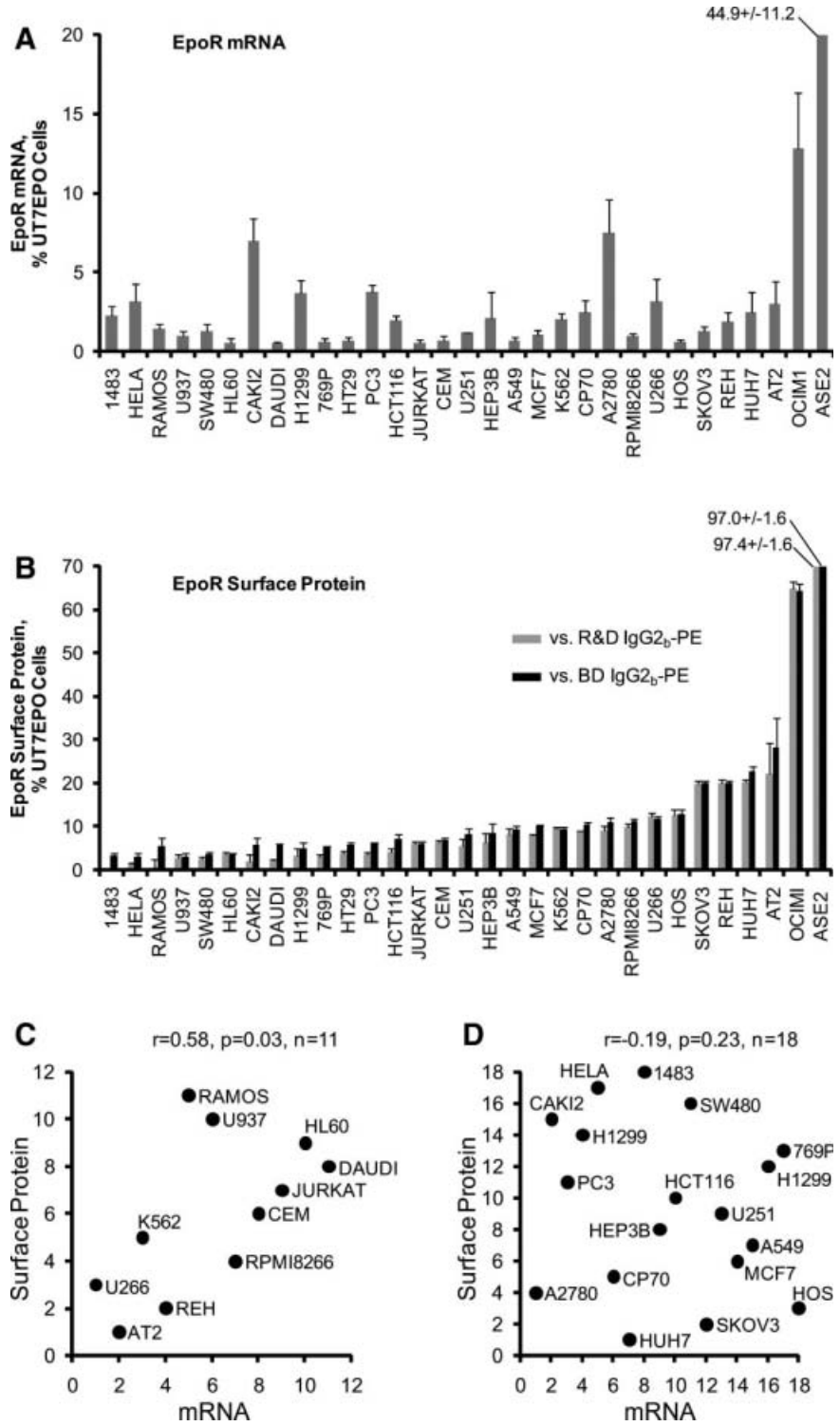

Figure 1. *EpoR* mRNA and surface protein levels in cancer cell lines. **(A):** *EpoR* mRNA levels were determined by quantitative reverse transcriptase polymerase chain reaction. The mRNA level of each cancer cell line is plotted relative to the level in control UT7EPO cells. Error bars represent the standard deviation of results obtained upon normalization to *Hmbs*, *Rplp0*, and *Hprt1*. **(B):** Differences between the mean fluorescent intensities of cells stained with a phycoerythrin (PE)-conjugated monoclonal anti-EpoR antibody versus each of two different PE-conjugated murine IgG_{2b}-PE isotype controls are depicted. Results are plotted relative to UT7EPO cells. Error bars depict standard deviations of the differences. **(C):** The rank order of mRNA and protein expression is plotted for all nonadherent cell lines, excluding the positive control lines UT7EPO, ASE2, and OCIM1. **(D)**: The rank order of mRNA and protein expression is plotted for all adherent cell lines. Spearman's rank order correlation coefficients are indexed above the graphs in **(C)** and **(D)**.

reports of Epo-dependent signal transduction in nonerythroid cells [9, 29].

Development of a Quantitative RT-PCR Assay for *EpoR* mRNA in Archival Tumor Samples

Most tumors from clinical trials are preserved as FFPE tissue. We found that immunohistochemistry with a specific antibody was not sufficiently sensitive to detect low-level EpoR protein in FFPE tumor cell lines (supporting information Fig. 2). We therefore tested whether *EpoR* mRNA could be accurately measured in FFPE tumors despite the RNA degradation that accompanies FFPE-processing [30]. Three independent assessments of *EpoR* mRNA levels from serial sections of 11 FFPE breast tumors demonstrated that our measurements were highly reproducible, and that *EpoR* mRNA levels varied as much as 34-fold (supporting information Fig. 3). To assess the validity of mRNA measurements from FFPE primary

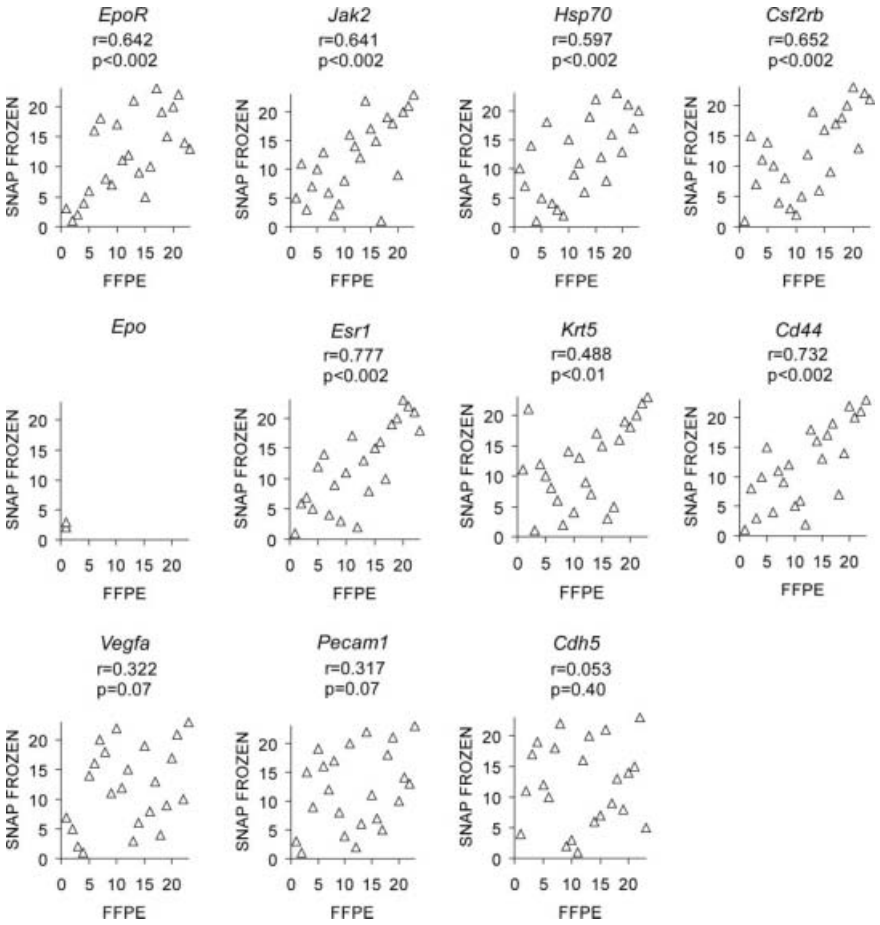

Figure 2. Analysis of concordance in mRNA measurements between snap frozen and formalin-fixed paraffin-embedded (FFPE) breast tumors. Twenty-three breast tumors are ranked for their level of mRNA expression of the indicated genes (normalized to *Hmbs* expression). Results using RNA extracted from snap frozen (y-axis) versus FFPE (x-axis) pieces of the same breast tumor are shown. Spearman's rank order correlation coefficients are indexed above each graph.

tumors, we compared expression levels for *EpoR* using 23 breast tumors which were divided and processed both as FFPE and snap-frozen tissue. Because the FFPE and snap-frozen samples represent different pieces of the same tumor, and snap-freezing preserves a higher degree of RNA integrity, this comparison allowed us to simultaneously assess whether RNA degradation influences the accuracy of our measurements as well as the uniformity with which *EpoR* is expressed across tumors. We also measured mRNA levels of *Jak2* and *Hsp70*, which participate in Epo signaling in erythroid cells [16, 21], *Csf2rb*, which has been suggested to enhance Epo signaling in nonerythroid cells [15], endothelial-associated genes (*Cdh5, Pecam1, Vegfa*), the squamous epithelial marker *Krt5* (especially relevant for head and neck cancer) [22], the putative cancer stem cell marker *Cd44* [23], and *Epo* itself. Significant correlations between FFPE and snap-frozen mRNA measurements were observed for *EpoR*, *Csf2rb*, *Jak2*, *Hsp70*, *Cd44, Krt5* and *Esr1* (estrogen receptor-1, used as a positive control) (Fig. 2). These findings suggest that single tumor sections can be used to gauge the overall expression levels of these markers. Expression levels varied over a wide range for each these genes (Fig. 3A). In contrast, *Vegfa*, *Cdh5*, and *Pecam1* were not significantly correlated, consistent with regional heterogeneity in tumor vascularity [31] whereas *Epo* was detected in too few FFPE tumors to permit calculation of a correlation coefficient.

Assessing *EpoR* mRNA Levels in Head and Neck Cancers from ENHANCE

We assayed *EpoR* mRNA levels in 136 archival FFPE head and neck tumors from ENHANCE, a subset of the 154 evaluated previously by immunohistochemistry using the C20 antibody [13]. Since most samples consisted of only a single microscope slide with minimal tissue, we employed a nonbiased target-specific cDNA preamplification method for all genes (supporting information Table 3). We included 7 candidate reference genes for normalization (*Hprt1*, *Ppia*, *Ipo8*, *Hmbs*, *Gapdh*, *Tfrc*, and *Rplp0*). These reference genes were included based on their stability among 16 candidates tested by the Genorm algorithm [27] in a panel of 8 breast cancers and 8 head and neck cancers (supporting information Table 2). Results for *Hprt1* were excluded because of high or no Ct values in many samples. For tumor samples with sufficient RNA (representing 123 different tumors), there were strong positive correlations in Ct values among all reference genes ($r \geq 0.88$ for all pairwise comparisons, $p < .001$).

We tested normalization of *EpoR* values with each of the reference genes and assessed the extent to which relative quantification values might be influenced by RNA abundance/integrity. Specifically, a phenomenon was reported by Cronin et al. in which greater age of FFPE blocks (and the lower RNA abundance/integrity) was associated with higher relative quantification values even after normalization [32]. This effect was attributed to differential degradation of target versus endogenous control gene transcripts and was reduced by minimizing the size and range of target and control gene assay amplicon sizes. In our data set, higher reference gene Ct values (less RNA abundance/integrity) were indeed associated with higher relative *EpoR* quantification upon normalization to *Ppia* levels as evidenced by the strong positive correlation between *Ppia* Ct values and normalized *EpoR* relative quantification values (supporting information Fig. 4A). Similar results were obtained for

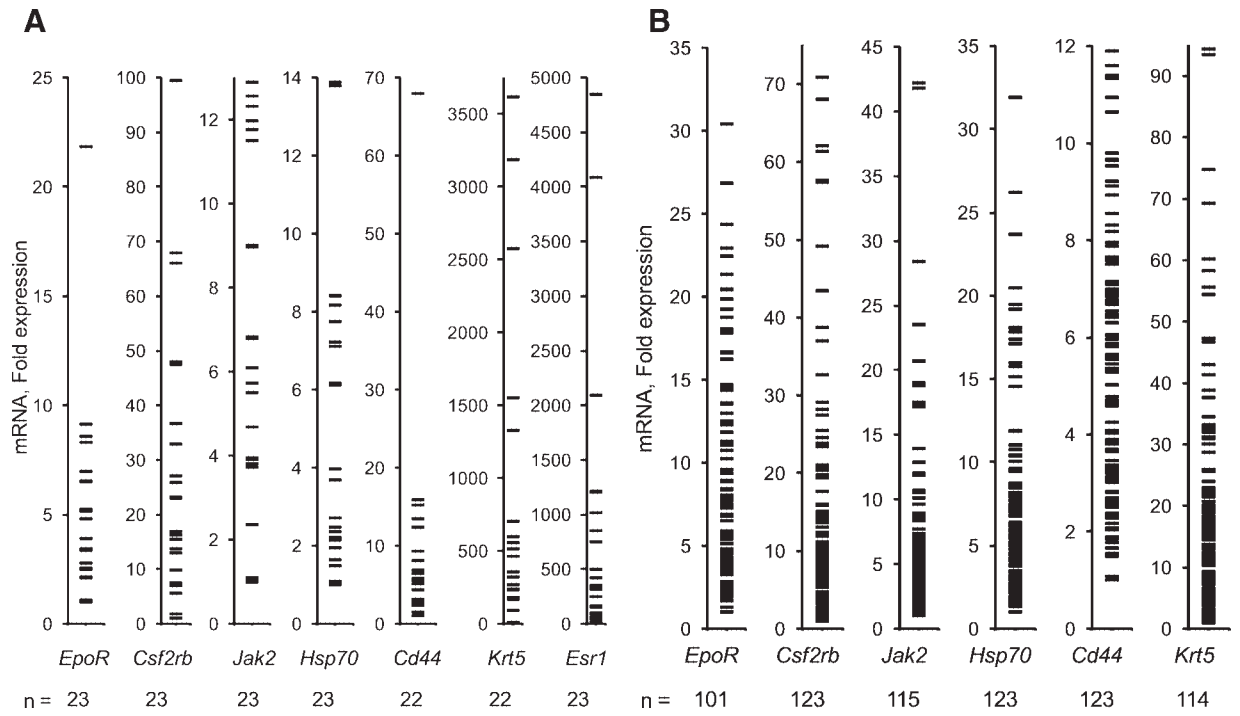

Figure 3. Range of mRNA levels in primary tumors. Relative quantification values are shown as fold differences, with the value from the lowest-expressing tumor assigned a value of 1. Numbers at the bottom of each graph indicate the number of tumors for which data were obtained. **(A)**: Results for breast tumors. **(B)**: Results for head and neck tumors from ENHANCE.

normalization with *Gapdh*, *Rplp0*, *Ipo8* or *Tfrc*. Whereas a similar pattern of normalized *EpoR* expression within subgroups of tumor samples with similar amounts of RNA abundance/integrity was observed, the overall systematic effect of RNA abundance/integrity on normalization would have precluded comparisons among all patients. Consistent with Cronin et al., this systematic effect was alleviated upon normalization to *Hmbs*, which had the shortest amplicon size among all reference gene assays tested (64 bp) (supporting information Fig. 4B). Five samples with low RNA abundance/integrity produced relative quantification values greater than mean + 1 standard deviation and were omitted. For 30 tumors, *EpoR* mRNA levels could not be determined relative to other samples because of undetermined endogenous control gene (*Hmbs*) and/or *EpoR* Ct values. Among the remaining 101 tumors, we observed a >30-fold range of *EpoR* mRNA, and expression levels also varied widely for the other genes examined (Fig. 3B).

mRNA Levels and Locoregional Progression-Free Survival

We evaluated LPFS within each resection stratum for patients with tumors expressing above- versus below-median levels of each marker (determined separately for each stratum). Significant associations between transcript level, Epo treatment and adverse outcome were observed only in the no resection stratum ($n = 28$) (Table 1). Significantly poorer LPFS was observed for Epo-treated subjects with above-median but not below-median levels of *EpoR* (Fig. 4A) or *Jak2* (Fig. 4B) (*EpoR*: above-median $p = .02$, $n = 14$, below-median $p = .8$, $n = 14$; *Jak2*: above-median $p = .04$, $n = 17$, below-median $p = .34$, $n = 18$). In addition, we found a significant association between Epo treatment, poor outcome, and below-median but not above-median levels of *Hsp70* family members in aggregate (Fig. 4C) (*Hsp70* below-median $p = .01$, $n = 20$, above-median $p = .38$, $n = 19$) and individually (supporting information Table 4). The significance of these associations was not further increased by dichotomizing mRNA at higher thresholds (that is, highest 10% vs. the rest) (not shown). Combinations of above-median *EpoR*, above-median *Jak2*, and below-median *Hsp70* did not increase the significance of the association between treatment assignment (Epo versus placebo) and LPFS compared to each marker individually (not shown).

We also compared Epo-treated patients with above-median *EpoR* expression to Epo-treated patients with below-median expression in the no resection stratum (see bracket below the graphs in Fig. 4A). A trend toward worse LPFS in Epo-treated patients with above-median levels of tumor *EpoR* mRNA was not significantly different from patients whose tumors expressed below-median *EpoR* mRNA ($p = .13$, $n = 11$). Analogous comparisons for *Jak2* and *Hsp70* mRNA levels also showed no significant differences among Epo-treated patients.

Relationship to Endogenous Erythropoietin Levels

If exogenous Epo can stimulate tumor progression, endogenous Epo might also stimulate tumor progression. The ENHANCE study documented single-time-point pretreatment serum Epo levels [3], and we obtained these results for 147 of the 154 patients reported previously [13] from the trial sponsor. Confining analyses to subjects enrolled in the placebo group, we did not find an association between LPFS and baseline hemoglobin level or LPFS and baseline serum Epo level (not shown). Additionally, baseline hemoglobin levels and serum Epo levels did not correlate (not shown). Finally, we tested whether elevated endogenous Epo levels were associated with LPFS in subjects with above-median versus below-median levels of *EpoR*, *Jak2* or *Hsp70* mRNA. Because of the small number of patients available, we combined the incomplete and no resection strata into a new category called "residual tumor." However, there was no association between

Table 1. Analysis of exogenous erythropoietin administration and locoregional progression-free survival by mRNA marker status[a]

mRNA Marker/Resection Stratum	Below Median[b] Marker Value			Above Median[b] Marker Value		
	Number of Patients		Log Rank	Number of Patients		Log Rank
	Epo	Placebo	*p* Value[c]	Epo	Placebo	*p* Value[c]
EpoR						
All patients	24	27	0.22	23	27	0.31
Complete	11	13	0.90	14	10	0.47
Incomplete	5	8	0.17	6	6	0.82
No resection	7	7	0.80	4[d]	10[d]	0.02[d]
Jak2						
All patients	29	29	0.65	26	31	0.15
Complete	14	14	0.12	14	13	0.15
Incomplete	7	6	0.33	4	8	0.71
No resection	8	10	0.34	8[d]	9[d]	0.04[d]
Hsp70[e]						
All Patients	31	31	0.98	29	32	0.40
Complete	15	15	0.26	15	14	0.74
Incomplete	6	7	0.79	5	7	0.16
No resection	9[d]	11[d]	0.01[d]	10	9	0.38
Csf2rb						
All patients	24	29	0.55	26	27	0.83
Complete	9	17	0.57	17	8	0.34
Incomplete	5	8	0.78	6	6	0.18
No resection	9	6	0.35	4	11	0.26
Cd44						
All Patients	30	32	0.73	30	31	0.50
Complete	16	14	0.70	14	15	0.96
Incomplete	5	8	0.52	6	6	0.89
No resection	9	11	0.12	10	9	0.10
Krt5						
All Patients	30	27	0.84	27	30	0.28
Complete	17	10	0.86	10	16	0.61
Incomplete	4	9	0.89	7	5	0.65
No resection	10	8	0.26	9	9	0.08

[a]Stratification was above versus below/equal to the median.
[b]The median was calculated separately for all patients and within each resection stratum.
[c]The *p* value is two sided and is based on the log rank test to compare differences in Kaplan Meier distributions in response to Epo versus placebo.
[d]Groups with significant adverse effects of Epo.
[e]*Hsp70* mRNA represents the cumulative expression of all 8 family members. Results for individual family members are presented in Supporting information Table 4.

endogenous Epo level and LPFS based on tumor *EpoR*, *Jak2* or *Hsp70* mRNA levels (supporting information Table 5).

Correlations with C20 Staining

The tumors we evaluated were among the 154 previously characterized using the C20 antibody [13], which was raised against a human EpoR sequence but cross-reacts with other proteins, including Hsp70 family members [14]. Notably, we found no significant correlation between C20 status and *EpoR* mRNA ($r = -0.11$, $p = .26$, $n = 100$), or *Hsp70* family member mRNA in aggregate ($r = 0.06$, $p = .54$, $n = 122$) or individually (supporting information Fig. 5).

Discussion

As one of the most prominent drugs in oncology, the unexpected association between Epo and increased cancer death rates has created concern and uncertainty. One of the central questions is whether Epo induced "off-target" signaling in tumors or tumor blood vessels can hasten cancer progression. Preclinical models may bring insight to this issue, however definitive answers can only come from studies in humans. The substantial challenge presented by the very low level of EpoR present in nonerythroid cells is counter-balanced in part by an extensive body of literature surrounding Epo signaling in erythroid cells. Using this knowledge, we developed methods to characterize human tumors for their potential competency to respond to Epo. Since existing reagents for detecting EpoR protein in tumor sections are insufficiently sensitive and specific [14], we measured mRNA. Laying the foundation for this effort, we show that *EpoR* mRNA levels can be estimated despite the extensive RNA degradation that accompanies FFPE processing and, barring results of adherent cell lines that are difficult to accurately assess by flow cytometry, we show that *EpoR* mRNA levels appear to reasonably estimate levels of EpoR cell surface protein.

We found a >30-fold range of *EpoR* mRNA across a series of breast cancers and head and neck cancers. This finding does not necessarily contradict a previous study documenting the lack of significant differences in *EpoR* mRNA levels between tumors and normal tissues [33]. A preferential susceptibility to Epo-induced signaling in malignant versus normal tissue is not a prerequisite for a direct effect of Epo on tumors, for example, a heritable basis for variation in EpoR expression levels has been proposed in swine [34]. Measurements of total tumor *EpoR* mRNA levels cannot distinguish

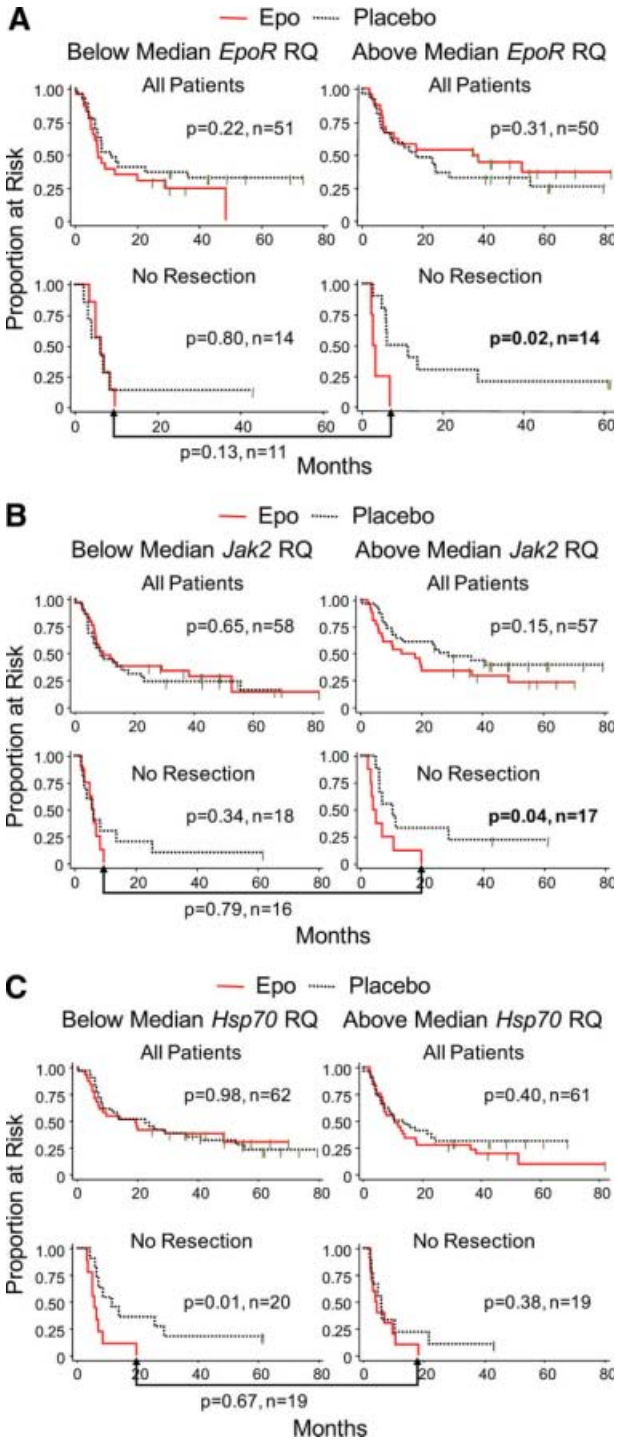

Figure 4. Effects of exogenous Epo on LPFS with stratification by mRNA status. Outcomes in response to Epo versus placebo are shown in Kaplan-Meier plots. The log-rank *p* value is two sided. Comparisons of outcomes of patients randomized to Epo are indexed below the brackets. **(A):** Results for *EpoR*. **(B):** Results for *Jak2*. **(C):** Results for *Hsp70*. *Hsp70* mRNA measurements reflect the cumulative expression of all eight family members.

the cellular origin of the *EpoR* transcript, and the extent to which the various cell types within tumors might respond to Epo and contribute to tumor progression remains undetermined. Future efforts directed at laser capture microdissection of various cell types from tumor samples will help resolve this issue. Wide variations in *Jak2*, *Hsp70*, and *Csf2rb* mRNA levels were also found across our series of breast cancers and head and neck cancers.

We used these tools to hunt for an association between Epo exposure, a tumor's inferred competency to respond to Epo based on mRNA levels of Epo-associated signaling molecules and patient outcomes. The ideal testing grounds for this effort are the archival tumors of patients who were randomized in clinical trials of Epo versus placebo and whose outcomes are known. For this first study, we examined available tumors from ENHANCE [3]. Above-median levels of mRNA for *EpoR*, and its tethered signaling intermediate *Jak2*, emerged as candidate predictors of reduced LPFS in unresected patients treated with Epo compared to placebo. In contrast, we found no significant association between *Csf2rb, Krt5, or Cd44* mRNA levels and outcome. Since Hsp70 mediates Epo signaling in erythroid cells [21] and is detected by the C20 antibody [14], we predicted correlations between *Hsp70* mRNA levels, Epo and LPFS analogous to those observed with *EpoR* and *Jak2*. To the contrary, we found a strong association between below-median levels of all *Hsp70* family members, Epo treatment and poor outcome. Importantly, we did not find a correlation between *EpoR* mRNA levels and prior staining of the same tumors using the C20 antibody [13]. These findings are consistent with the interpretation that the C20 staining does not correlate with EpoR expression [14]. Confirmation of C20 staining as a predictor of tumor susceptibility to Epo may lead to the identification of other proteins involved in Epo responsiveness.

Whether these tentative associations are reflective of underlying tumor biology is unknown, and the interpretation of our findings must be tempered by several limitations. First, contrasting the disparate outcomes of subjects randomized to Epo versus placebo, mRNA levels were not associated with significant differences in LPFS when restricting our analysis to patients in the Epo-treated group (see the brackets beneath the Kaplan Meier plots in Fig. 4). Thus, differences in LPFS associated with above- versus below-median mRNA levels cannot be accounted for entirely by differences in outcomes in response to Epo. Similarly, we found no association between these markers and adverse outcome in the presence of elevated levels of endogenous Epo in patients enrolled in the placebo arm of ENHANCE. Second, the statistically significant associations that we observed were confined to patients with unresected tumors, and did not extend to patients with incomplete or complete resection of their tumors. This discrepancy might be explained by the prediction that, absent resection, a larger amount of tumor would be available for Epo stimulation. In support of this interpretation, the original ENHANCE trial also did not find an association between Epo treatment and worse outcomes in patients with completely resected tumors [3]. Third, the small number of patients with unresected tumors also precludes further stratification to adjust for potential baseline imbalances or confounding clinical characteristics. Fourth, in view of the exploratory nature of our hypothesis, we did not adjust for multiple comparisons. This increases the likelihood that the observed significant *p* values represent false positives [35]. Most importantly, our findings are constrained by lack of access to additional tumors from other Phase III clinical trials of Epo. Because these were large multicenter trials lacking centralized tumor repositories, obstacles to obtaining these tumors likely can only be surmounted by the trial sponsors. Tapping this little explored resource using the methods described here may bring new insight to Epo and cancer progression.

ACKNOWLEDGMENTS

We thank Roche for providing baseline serum Epo levels for patients enrolled in ENHANCE, Janet Rowley for AT-2 cells, and Chugai Pharmaceutical for ASE2 cells. We gratefully acknowledge Erica Jonlin for assistance with human subjects issues, Elizabeth Wayner for antibody development, Julie Randolph Habecker for immunohistochemistry, Margaret Pepe and Ruth Etzioni for statistical advice, Kerstin Edlefsen and Keith Loeb for assistance with pathology specimens, Kathy O'Briant, Leah Sabacan and J. David Beatty for breast cancer samples and helpful discussions, and Whitney Hitchcock, Philamer Calses, and Marianna Rudakova for assistance with flow cytometry analyses. This work was supported by the National Institutes of Health [1R01CA135357 to C.A.B.]; and the Department of Defense [CDMRP, DAMD17-02-0691 to N.U.].

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

C. Anthony Blau owned stock in and served as an officer or member of the board for CellNexus, LLC.

REFERENCES

1 Caro JJ, Salas M, Ward A et al. Anemia as an independent prognostic factor for survival in patients with cancer: A systemic, quantitative review. Cancer 2001;91:2214–2221.
2 Bennett CL, Silver SM, Djulbegovic B et al. Venous thromboembolism and mortality associated with recombinant erythropoietin and darbepoetin administration for the treatment of cancer-associated anemia. JAMA 2008;299:914–924.
3 Henke M, Laszig R, Rübe C et al. Erythropoietin to treat head and neck cancer patients with anaemia undergoing radiotherapy: Randomised, double-blind, placebo-controlled trial. Lancet 2003;362:1255–1260.
4 Leyland-Jones B, Semiglazov V, Pawlicki M et al. Maintaining normal hemoglobin levels with epoetin alfa in mainly nonanemic patients with metastatic breast cancer receiving first-line chemotherapy: a survival study. J Clin Oncol 2005;23:5960–5972.
5 Wright JR, Ung YC, Julian JA et al. Randomized, double-blind, placebo-controlled trial of erythropoietin in non-small-cell lung cancer with disease-related anemia. J Clin Oncol 2007;25:1027–1032.
6 Smith RE Jr., Aapro MS, Ludwig H et al. Darbepoetin alfa for the treatment of anemia in patients with active cancer not receiving chemotherapy or radiotherapy: Results of a phase III, multicenter, randomized, double-blind, placebo-controlled study. J Clin Oncol 2008;26:1040–1050.
7 Oncologic Drugs Advisory Committee meeting briefing materials at http://www.fda.gov/ohrms/dockets/ac/08/briefing/2008–4345b2–01-FDA.pdf. Accessed March 13, 2008.
8 Blau CA. Erythropoietin in cancer: Presumption of innocence? Stem Cells 2007;25:2094–2097.
9 Arcasoy MO. Erythropoiesis-stimulating agent use in cancer: Preclinical and clinical perspectives. Clin Cancer Res 2008;14:4685–4690.
10 Acs G, Acs P, Beckwith SM et al. Erythropoietin and erythropoietin receptor expression in human cancer. Cancer Res 2001;61:3561–3565.
11 Sinclair AM, Todd MD, Forsythe K et al. Expression and function of erythropoietin receptors in tumors: Implications for the use of erythropoiesis-stimulating agents in cancer patients. Cancer 2007;110:477–488.
12 Um M, Gross AW, Lodish HF. A "classical" homodimeric erythropoietin receptor is essential for the antiapoptotic effects of erythropoietin on differentiated neuroblastoma SH-SY5Y and pheochromocytoma PC-12 cells. Cell Signal 2007;19:634–645.
13 Henke M, Mattern D, Pepe M et al. Do erythropoietin receptors on cancer cells explain unexpected clinical findings? J Clin Oncol 2006;24:4708–4713.
14 Elliott S, Busse L, Bass MB et al. Anti-Epo receptor antibodies do not predict Epo receptor expression. Blood 2006;107:1892–1895.
15 Brines M, Grasso G, Fiordaliso F et al. Erythropoietin mediates tissue protection through an erythropoietin and common beta-subunit heteroreceptor. Proc Natl Acad Sci U S A 2004;101:14907–14912.
16 Ihle JN, Gilliland DG. Jak2: normal function and role in hematopoietic disorders. Curr Opin Genet Dev 2007;17:8–14.
17 Huang LJ, Constantinescu SN, Lodish HF. The N-terminal domain of Janus kinase 2 is required for Golgi processing and cell surface expression of erythropoietin receptor. Mol Cell 2001;8:1327–1338.
18 Digicaylioglu M, Lipton SA. Erythropoietin-mediated neuroprotection involves cross-talk between Jak2 and NF-kappaB signalling cascades. Nature 2001;412:641–647.
19 Morano KA. New tricks for an old dog: The evolving world of Hsp70. Ann N Y Acad Sci 2007;1113:1–14.
20 Sherman M, Multhoff G. Heat shock proteins in cancer. Ann N Y Acad Sci 2007;1113:192–201.
21 Ribeil JA, Zermati Y, Vandekerckhove J et al. HSP70 regulates erythropoiesis by preventing caspase-3-mediated cleavage of GATA-1. Nature 2007;445:102–105.
22 Becker MT, Shores CG, Yu KK et al. Molecular assay to detect metastatic head and neck squamous cell carcinoma. Arch Otolaryngol Head Neck Surg 2004;130:21–27.
23 Prince ME, Sivanandan R, Kaczorowski A et al. Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma. Proc Natl Acad Sci U S A 2007;104:973–978.
24 Phillips TM, Kim K, Vlashi E et al. Effects of recombinant erythropoietin on breast cancer-initiating cells. Neoplasia 2007;9:1122–1129.
25 Palacios R, Steinmetz M. Il-3-dependent mouse clones that express B-220 surface antigen, contain Ig genes in germ-line configuration, and generate B lymphocytes in vivo. Cell 1985;41:727–734.
26 Inoue Y, Tsushima H, Ando K et al. Chemokine expression in human erythroid leukemia cell line AS-E2: Macrophage inflammatory protein-3alpha/CCL20 is induced by inflammatory cytokines. Exp Hematol 2006;34:19–26.
27 Vandesompele J, De Preter K, Pattyn F et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002;3:RESEARCH0034.
28 Paul I, Lappin TR, Maxwell P et al. Pre-operative plasma erythropoietin concentration and survival following surgery for non-small cell lung cancer. Lung Cancer 2006;51:329–334.
29 Fu P, Jiang X, Arcasoy MO. Constitutively active erythropoietin receptor expression in breast cancer cells promotes cellular proliferation and migration through a MAP-kinase dependent pathway. Biochem Biophys Res Commun 2009;379:696–701.
30 Antonov J, Goldstein DR, Oberli A et al. Reliable gene expression measurements from degraded RNA by quantitative real-time PCR depend on short amplicons and a proper normalization. Lab Invest 2005;85:1040–1050.
31 Grizzi F, Colombo P, Taverna G et al. Geometry of human vascular system: is it an obstacle for quantifying antiangiogenic therapies? Appl Immunohistochem Mol Morphol 2007;15:134–139.
32 Cronin M, Pho M, Dutta D et al. Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. Am J Pathol 2004;164:35–42.
33 Sinclair AM, Rogers N, Busse L et al. Erythropoietin receptor transcription is neither elevated nor predictive of surface expression in human tumour cells. Br J Cancer 2008;98:1059–1067.
34 Vallet JL, Freking BA, Leymaster KA et al. Allelic variation in the erythropoietin receptor gene is associated with uterine capacity and litter size in swine. Anim Genet 2005;36:97–103.
35 Wang R, Lagakos SW, Ware JH et al. Statistics in medicine–reporting of subgroup analyses in clinical trials. N Engl J Med 2007;357:2189–2194.


See www.StemCells.com for supporting information available online.

List of Personnel Receiving Pay (Oct. 2008-Sep. 2009)

Garnet Anderson, PhD
(Marin) Robyn Andersen, PhD
Charles Drescher, MD
Allen Gown, MD
Martin McIntosh, PhD
Michèl Schummer, PhD
Nicole Urban, ScD

Kelsey Afdem	(Program Assistant)
Kiea Arispe	(Program Assistant)
Lindsay Bergan	(Research Technician)
Archana Kampani	(Research Technician)
Christine Linders	(Program Assistant)
Jeff Brown	(Programmer)
Robin Forrest	(Research Technician)
Becky Gage	(Office Worker)
Barbara Garvik	(Research Technician)
Ann Green	(Research Technician)
Lisa Grimm	(Program Assistant)
Kathy O'Briant	(Program Manager)
Michael O'Donnell	(Programmer)
Jeffrey Peterson	(Programmer)
Kristi Schurman	(Project Coordinator)
Carole Shaw	(Database Manager)
Jason Thorpe	(Statistical Research Associate)
Kate Watabayashi	(Project Coordinator)
Lauren Zimmerman	(Program Assistant)